

Honghai Liu  
Han Ding  
Zhenhua Xiong  
Xiangyang Zhu (Eds.)

LNAI 6424

# Intelligent Robotics and Applications

Third International Conference, ICIRA 2010  
Shanghai, China, November 2010  
Proceedings, Part I

1  
Part I

 Springer

Lecture Notes in Artificial Intelligence

6424

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Honghai Liu Han Ding Zhenhua Xiong  
Xiangyang Zhu (Eds.)

# Intelligent Robotics and Applications

Third International Conference, ICIRA 2010  
Shanghai, China, November 10-12, 2010  
Proceedings, Part I

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada  
Jörg Siekmann, University of Saarland, Saarbrücken, Germany  
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

Honghai Liu  
The University of Portsmouth, School of Creative Technologies  
Portsmouth PO1 2DJ, UK  
E-mail: honghai.liu@port.ac.uk

Han Ding  
Shanghai Jiao Tong University, Robotics Institute, Shanghai 200240, China  
E-mail: hding@sjtu.edu.cn

Zhenhua Xiong  
Shanghai Jiao Tong University, Robotics Institute, Shanghai 200240, China  
E-mail: mexiong@sjtu.edu.cn

Xiangyang Zhu  
Shanghai Jiao Tong University, Robotics Institute, Shanghai 200240, China  
E-mail: mexyzhu@sjtu.edu.cn

Library of Congress Control Number: 2010936828

CR Subject Classification (1998): I.4, I.5, I.2, I.2.10, H.4, C.2

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743

ISBN-10 3-642-16583-4 Springer Berlin Heidelberg New York  
ISBN-13 978-3-642-16583-2 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2010  
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India  
Printed on acid-free paper 06/3180



# Preface

The market demand for skills, knowledge and adaptability have positioned robotics to be an important field in both engineering and science. One of the most highly visible applications of robotics has been the robotic automation of many industrial tasks in factories. In the future, a new era will come in which we will see a greater success for robotics in non-industrial environments. In order to anticipate a wider deployment of intelligent and autonomous robots for tasks such as manufacturing, healthcare, entertainment, search and rescue, surveillance, exploration, and security missions, it is essential to push the frontier of robotics into a new dimension, one in which motion and intelligence play equally important roles.

The 2010 International Conference on Intelligent Robotics and Applications (ICIRA 2010) was held in Shanghai, China, November 10–12, 2010. The theme of the conference was “Robotics Harmonizing Life,” a theme that reflects the ever-growing interest in research, development and applications in the dynamic and exciting areas of intelligent robotics. These volumes of Springer’s *Lecture Notes in Artificial Intelligence* and *Lecture Notes in Computer Science* contain 140 high-quality papers, which were selected at least for the papers in general sessions, with a 62% acceptance rate. Traditionally, ICIRA 2010 holds a series of plenary talks, and we were fortunate to have two such keynote speakers who shared their expertise with us in diverse topic areas spanning the rang of intelligent robotics and application activities.

We would like to thank the International Advisory Committee for their guidance and advice. We are also especially grateful to the Program Committee members for their rigorous and efficient reviewing of the submitted papers, and the Organizing Committee for their enormous efforts and excellent work. In addition, we greatly appreciate the Ministry of Education, National Natural Science Foundation of China, Science and Technology Commission of Shanghai Municipality, Shanghai Jiao Tong University, State Key Laboratory of Mechanical System and Vibration, State Key Lab of Digital Manufacturing Equipment & Technology, and Springer for their support.

It has been our pleasure to organize the 2010 International Conference on Intelligent Robotics and Applications. We hope it is your pleasure to read these selected papers by some of the highest-quality researchers and practitioners in the world of intelligent robotics. We also hope that they provide inspiration for your own research and help advance the state of the art in the multidisciplinary field of robotics and its application.

August 2010

Xiangyang Zhu  
Han Ding  
Zhenhua Xiong  
Honghai Liu

# Conference Organization

## International Advisory Committee

Tamio Arai	University of Tokyo, Japan
Hegao Cai	Harbin Institute of Technology, China
Toshio Fukuda	Nagoya University, Japan
Huosheng Hu	Essex University, UK
Oussama Khatib	Stanford University, USA
Jurgen Leopold	Fraunhofer IWU, Germany
Ming Li	National Natural Science Foundation of China
Peter Luh	Connecticut University, USA
Jun Ni	University of Michigan, USA
Nikhil R. Pal	Indian Statistical Institute, India
Grigory Panovko	Russian Academy of Sciences, Russia
Shigeki Sugano	Waseda University, Japan
Michael Wang	Chinese University of Hong Kong, China
Bogdan M. Wilamowski	Auburn University, USA
Ming Xie	Nanyang Technological University, Singapore
Lotfi Zadeh	University of California, Berkeley, USA

## General Chairs

Zhongqin Lin	Shanghai Jiao Tong University, China
Youlun Xiong	Huazhong University of Science and Technology, China

## Program Chairs

Xiangyang Zhu	Shanghai Jiao Tong University, China
Honghai Liu	University of Portsmouth, UK

## Organizing Committee Chairs

Han Ding	Shanghai Jiao Tong University, China
Zexiang Li	Hong Kong University of Science & Technology, China
Hong Liu	Harbin Institute of Technology, China
Guang Meng	China Aerospace Science & Technology Co., China
Guobiao Wang	National Natural Science Foundation of China
Tianmiao Wang	Beihang University, China

## Area Chairs

H. Levent Akin	Bogazici University, Turkey
Kaspar Althoefer	King's College London, UK
Shuzhi Sam Ge	National University of Singapore, Singapore
Qiang Huang	Beijing Institute of Technology, China
Naoyuki Kubota	Tokyo Metropolitan University, Japan
Chunyi Su	Concordia University, Canada
Jun Wang	Chinese University of Hong Kong, China
Diedrich Wolter	Bremen University, Germany
Bin Yao	Purdue University, USA
Jie Zhao	Harbin Institute of Technology, China

## Tutorial/Workshop Chairs

Ryad Chellali	Institute of Italy Technologies, Italy
Qixin Cao	Shanghai Jiao Tong University, China

## Poster Session Chairs

James Hu	Kumamoto University, Japan
Min Lei	Shanghai Jiao Tong University, China

## Publicity Chairs

Mohamed Kamel	University of Waterloo, Canada
Yun Luo	Shanghai Jiao Tong University, China

## Publication Chairs

Rainer Palm	Orebro University, Sweden
Yanzheng Zhao	Shanghai Jiao Tong University, China

## Financial Chair

Zhenhua Xiong	Shanghai Jiao Tong University, China
---------------	--------------------------------------

## Registration Chair

Dingguo Zhang	Shanghai Jiao Tong University, China
---------------	--------------------------------------

## Local Chair

Lifeng Xi	Shanghai Jiao Tong University, China
-----------	--------------------------------------

## General Affairs and Secretariat

Xinjun Sheng

Shanghai Jiao Tong University, China

## Acknowledgment

Ministry of Education, China

National Natural Science Foundation of China

Science and Technology Commission of Shanghai Municipality, China

Shanghai Jiao Tong University, China

State Key Laboratory of Mechanical System and Vibration, China

State Key Lab of Digital Manufacturing Equipment & Technology, China

## List of Reviewers

We would like to acknowledge the support of the following people who contributed to peer review of articles from ICIRA 2010.

Absi Alfaro	Wen-Jie Chen	Nicola Ferrier
Cihan Acar	Xiao-Peng Chen	Kevin Fite
Mojtaba Ahmadi	You-Hua Chen	Zhuang Fu
Eric L Akers	Sonia Chernova	Antonios Gasteratos
H. Levent Akin	Yu-Zheng Chong	Shu-Zhi Sam Ge
Berkant Akin	Henrik Christensen	Akio Gofuku
Mohamed Al Marzouqi	Dimitrios Chrysostomou	Hu Gong
Redwan Alqasemi	Sadek Crisóstomo	Darwin Gouwanda
Kaspar Althoefer	Xavier Cufi	Dong-Bing Gu
Farshid Amirabdollahian	Ravinder Dahiya	En-Guang Guan
Cecilio Angulo	Konstantinos Dalamagkidis	Li-Wen Guan
Panagiotis Artemiadis	Greyson Daugherty	Lei Guo
Fernando Auat Cheein	Raoul de Charette	Wei-Zhong Guo
Joonbum Bae	Joris De Schutter	Alvaro Gutierrez
Feng Bai	Liyanage De Silva	Norihiro Hagita
Malek Baklouti	Chandratilak	Hassan Haleh
Sven Behnke	Hua Deng	Tatsuya Harada
Nicola Bellotto	Jean-Yves Didier	Masaki Hasegawa
Roger Bostelman	Ming Ding	Kenji Hashimoto
Magdalena Bugajska	Hao Ding	Mitsuhiro Hayashibe
Darius Burschka	Can Ulas Dogruer	Vincent Hayward
Li-Xin Cao	Zhen-Chun Du	H. Hellendoorn
Jiang-Tao Cao	Damien Eynard	Patrick Henaff
Qi-Xin Cao	Bao-Jie Fan	Sophie Hennequin
Zhi-Qiang Cao	Hui-Jie Fan	Rafael Herrejon
Barbara Caputo	Janabi-Sharifi Farrokh	Bernard Voon Ee How
Guillaume Caron	Ying Feng	Ming-Hui Hu
Dian-Sheng Chen	Manuel Fernandez-Carmona	Chuan-Zhen Huang

Yong-An Huang	Yun Luo	Ching-Hua Ting
Cong-Hui Huang	Xiang Luo	Michael John Tribou
Atsushi Imiya	Jie Ma	Ching-Chih Tsai
Xiao-Fei Ji	Li Ma	Cristina Urdiales
Li Jiang	Dominic Maestas	Akira Utsumi
Zhao-Jie Ju	Elmar Mair	Kalyana Veluvolu
Agung Julius	Adrian Martin	Qi-Yuan Wang
Takahiro Kagawa	Emanuele Menegatti	Furui Wang
Chen-Yu Kai	Krzysztof Mianowski	Zongyao Wang
Shun-Ichi Kaneko	Huasong Min	Zhenyan Wang
Kikuhito Kawasue	Xie Ming	Hao Wang
Erdal Kayacan	Seyed Mohamed Buhari	Guowu Wei
Chong-Hui Kim	Mohamed Ismail	Ang Wei Tech
Eun-Young Kim	Hyungpil Moon	Myra Wilson
Sang-Ho Kim	Behzad Moslehi	Hong-Tao Wu
Chyon-Hae Kim	Josiah L. Munda	Xiao-Jun Wu
Won-Jong Kim	Hyun Myung	Chao Wu
Denis Klimentjew	Jawahar N.	Xian-Bo Xiang
Akio Kosaka	Hiroyuki Nakamoto	Cai-Hua Xiong
Naoyuki Kubota	Duc Dung Nguyen	Ji-Jie Xu
Ki-Ju Lee	Hirotsuka Osawa	Xi-Peng Xu
Bengt Lennartson	Mohammadreza Asghari	Xin Xu
Ales Leonardis	Oskoei	Zhao-Hong Xu
Gang Li	Nuno Otero	Bing Xu
Qin-Chuan Li	Chee Khiang Pang	Seiji Yamada
You-Fu Li	gaoliang Peng	Yong-Hua Yan
Yang-Min Li	Fernando Pereira	Wen-Yu Yang
Zhi-Jun Li	Anton Satria Prabuono	Yue-Hong Yin
Hai-Zhou Li	Flavio Prieto	Ding-Guo Zhang
Bin Li	Yue Qiu	Jian-Guo Zhang
Geng Liang	M. Jayedur Rashid	Wen-Zeng Zhang
Yong-Cheng Lin	Florian Raudies	Xu-Guang Zhang
Chen-Liang Liu	Jorge Rivera-Rovelo	Yu-Ru Zhang
Hai-Ning Liu	Debanik Roy	Xian-Min Zhang
Hong-Hai Liu	Alexander Schmitz	Yu-Nong Zhang
Jin-Dong Liu	Desire Sidibe	Yan-jiang Zhao
Jia Liu	Ponnambalam Sivalinga	Yan-Zheng Zhao
Bing-Bing Liu	Kai-Tai Song	Yong Zhao
Wei Liu	Chun-Yi Su	Hui-Yu Zhou
Ping Liu	Anan Suebsomran	Li-Min Zhu
Chao Liu	Yue-Gang Tan	Chi Zhu
Xin-Jun Liu	Da-Wei Tan	Chun-Gang Zhuang
Yong-Huai Liu	Hideyuki Tanaka	Jun Zou
Benny Lo	Bo Tao	Wei Zou
Yi Lu	Michael Thomason	

# Table of Contents – Part I

## Bionic Dexterous Hand

Cartesian Impedance Control on Five-Finger Dexterous Robot Hand DLR-HIT II with Flexible Joint . . . . .	1
<i>Zhaopeng Chen, Neal Y. Liu, Minghe Jin, Shaowei Fan, and Hong Liu</i>	
Changeable Grasping Force Unit for Under-Actuated Gripper with Self-adaptive Grasping . . . . .	13
<i>Deyang Zhao and Wenzeng Zhang</i>	
Efficient Grasp Planning with Reachability Analysis . . . . .	26
<i>Zhixing Xue and Ruediger Dillmann</i>	
Development of an Anthropomorphic Prosthetic Hand for Man-Machine Interaction . . . . .	38
<i>Nan Li, Li Jiang, Dapeng Yang, Xinqing Wang, Shaowei Fan, and Hong Liu</i>	
Dexterous and Self-adaptive Under-Actuated Humanoid Robot Hand: GCUA Hand II . . . . .	47
<i>Demeng Che and Wenzeng Zhang</i>	
A Novel Coupled and Self-adaptive Under-Actuated Grasping Mode and the COSA-DTS Hand . . . . .	59
<i>Jie Sun, Wenzeng Zhang, and Haitao Sun</i>	
Empirical Copula Driven Hand Motion Recognition via Surface Electromyography Based Templates . . . . .	71
<i>Zhaojie Ju and Honghai Liu</i>	
A Continuous Control Scheme for Multifunctional Robotic Arm with Surface EMG Signal . . . . .	81
<i>Pinghua Hu, Shunchong Li, Xinpu Chen, Dingguo Zhang, and Xiangyang Zhu</i>	

## Modeling and Control of Systems Involving Hysteresis

On the Robust Control of Systems Preceded by Differential Equation-Based Hysteresis Nonlinearities . . . . .	92
<i>Ying Feng, Juan Du, and Chun-Yi Su</i>	

A Generalized Play Operator for Modeling and Compensation of Hysteresis Nonlinearities . . . . .	104
<i>Mohammad Al Janaideh, Omar Aljanaideh, and Subhash Rakheja</i>	
Hysteresis Modeling of Piezoelectric Actuators Using the Fuzzy System . . . . .	114
<i>Pengzhi Li, Guoying Gu, Leijie Lai, and Limin Zhu</i>	
On PSO Based Bouc-Wen Modeling for Piezoelectric Actuator . . . . .	125
<i>ZhenYan Wang and JianQin Mao</i>	
Operator-Based Robust Nonlinear Control for Ionic Polymer Metal Composite with Uncertainties and Hysteresis . . . . .	135
<i>Aihui Wang, Mingcong Deng, and Dongyun Wang</i>	
Modeling of Hysteresis Nonlinearity Based on Generalized Bouc-Wen Model for GMA . . . . .	147
<i>Yongxin Guo and Jianqin Mao</i>	
<b>Control System Modeling and Applications</b>	
Motion Dynamics Modelling of an Electric Wheel Robot . . . . .	159
<i>Xuezhu Wang, Xiangtao Zhuan, Guilin Zheng, and Zheng Chen</i>	
Sliding Mode Control of Hybrid Joints for Wheeled Mobile Manipulators with Markovian Switching and Bounded Torques . . . . .	171
<i>Lei Dai, Dan Huang, and Zhijun Li</i>	
Timestamp Based Predictive Robot Control System over Real-Time Industrial Ethernet . . . . .	183
<i>Huasong Min, Tianmiao Wang, Hongxing Wei, and Zhenglin Liu</i>	
Design and Experimental Research on a Fast Zero-Finding Algorithm Based on the Incremental Angle Encoder . . . . .	195
<i>Zhonghua Miao, Xuyong Wang, Chengliang Liu, and Xiaojin Zhu</i>	
Decentralized Synchronization and Output Tracking Control of Nondiffusively Coupled Complex Dynamical Networks . . . . .	204
<i>Gequn Liu, Xiaoming Xu, and Lei Liu</i>	
Manipulation of a Mobile Modular Manipulator Interacting with the Environment with the Assistance of Tactile Sensing Feedback . . . . .	214
<i>Jingguo Wang and Yangmin Li</i>	
An RTC-Based Intelligent Programming System for Service Robots . . . . .	226
<i>Masaru Adachi, Zhen Zhang, Peihua Chen, and Qixin Cao</i>	

Marine Engine State Monitoring System Using Distributed Precedence Queue Mechanism in CAN Networks . . . . .	237
<i>Hyun Lee, Dong kyu Yi, Jun seok Lee, Gye-do Park, and Jang Myung Lee</i>	

Modeling and Fault Tolerant Controller Design for PM Spherical Actuator . . . . .	246
<i>Zheng Li</i>	

## Intelligent Control and Their Applications

Artificial Neural Network to Predict the Surface Maximum Settlement by Shield Tunneling . . . . .	257
<i>Jinli Qiao, Jianqin Liu, Wei Guo, and Yitong Zhang</i>	

Applying Grey-Fuzzy-Fuzzy Rule to Machine Vision Based Unmanned Automatic Vehicles . . . . .	266
<i>Chin-Wen Chuang and Jau-Rong Li</i>	

AWARE: Autonomous Wireless Agent Robotic Exchange . . . . .	276
<i>Cory Q. Nguyen, Benny Leong, Eric T. Matson, Anthony Smith, and Juan P. Wachs</i>	

Application of Wearable Miniature Non-invasive Sensory System in Human Locomotion Using Soft Computing Algorithm . . . . .	288
<i>Murad Alaqtash, Huiying Yu, Richard Brower, Amr Abdelgawad, Eric Spier, and Thompson Sarkodie-Gyan</i>	

The Mathematical Model and Control Scheme of a Four-Legged Robot Based on GZ-I and Note Module . . . . .	300
<i>Yinfeng Fang, Houxiang Zhang, Xiuli Li, and S.Y. Chen</i>	

## Robot and Automation in Tunneling

Electromechanical Coupling Model for Cutterhead Driving System of Shield Machines . . . . .	310
<i>Jianzhong Sun, Ran Liu, Yaqin Luo, and Wei Sun</i>	

Dimension Optimization of an Orientation Fine-Tuning Manipulator for Segment Assembly Device . . . . .	318
<i>Chao Wu, Xin-Jun Liu, and Jinsong Wang</i>	

Rotation and Gate Movement Compound Control of Screw Conveyor Hydraulic System with Disturbance Compensation . . . . .	329
<i>Lintao Wang, Guofang Gong, Hu Shi, and Huaiyin Liu</i>	

Applications of the Overlapping Nodes Method in Simulating Crack-Expansion under the Action of TBM Cutter . . . . .	341
<i>Kui Zhang and Yi-min Xia</i>	



The Simulation Analysis of the TBM Disc Cutter Phase Angle . . . . .	348
<i>Kai Wang, Qing Tan, Yimin Xia, and Zijun Xu</i>	
The Design of Shield Tunnelling Machine Cutter Header's Cutting Performance Test Bed . . . . .	356
<i>Zongming Zhu, Yimin Xia, Dezhi Luo, and Tao Teng</i>	
The Effect of TBM Hydraulic Piping System Parameters on FSI Vibration . . . . .	363
<i>Jinghua Xie, Ke Tian, and Dawei Yang</i>	
Station Point Optimization of Total Station in Shield Tunneling Method Based on Set-Covering . . . . .	372
<i>Bolan Li, Xu Zhang, and Limin Zhu</i>	
Study on Earth Pressure Acting on Cutter Face Based on Nonlinear Optimization . . . . .	383
<i>Xiangtao Hu, Yong'an Huang, and Zhouping Yin</i>	
Incident Angle Measurement of Collimated Laser Based on the Pinhole Principle . . . . .	392
<i>Minghua Pan and Guoli Zhu</i>	
 <b>Robot Mechanism and Design</b>	
A Novel Design for the Self-reconfigurable Robot Module and Connection Mechanism . . . . .	400
<i>Enguang Guan, Weixin Yan, Dongsheng Jiang, Zhuang Fu, and Yanzheng Zhao</i>	
The Forward Kinematics Analysis of 6-3 Stewart Parallel Mechanisms . . . . .	409
<i>Shili Cheng, Hongtao Wu, Chaoqun Wang, Yu Yao, and Jianying Zhu</i>	
Portable Multi-axis CNC: A 3-CRU Decoupled Parallel Robotic Manipulator . . . . .	418
<i>Dan Zhang, Zhen Gao, and Jijie Qian</i>	
Carton Motion-Moment Diagram and Stiffness Characteristics . . . . .	430
<i>Guowu Wei, Ruirui Zhang, and Jian S. Dai</i>	
Flexible Foot Design for Biped Walking on Uneven Terrain . . . . .	442
<i>Chen Koon Toh, Ming Xie, Hejin Yang, Guoqing Zhang, Quoc Phuong Bui, and Bo Tian</i>	
4WD Skid-Steer Trajectory Control of a Rover with Spring-Based Suspension Analysis: Direct and Inverse Kinematic Parameters Solution . . . . .	453
<i>Edgar Martínez-García and Rafael Torres-Córdoba</i>	

Topology and Analysis of Three-Phalanx COSA Finger Based on Linkages for Humanoid Robot Hands . . . . .	465
<i>Deyang Zhao and Wenzeng Zhang</i>	
Design of a Bionic Saltatorial Leg for Jumping Mini Robot . . . . .	477
<i>Xin Fu, Fei Li, Weiting Liu, Cesare Stefanini, and Paolo Dario</i>	
Asymmetrical Prototype of a Five-Wheeled Robot and Maneuver Analysis . . . . .	488
<i>He Xu, Jinfeng Zhao, Dawei Tan, and Zhenyu Zhang</i>	
Development of Controllable Two-Way Self-locking Mechanism for Micro In-Pipe Robot . . . . .	499
<i>Yong Xue, Jian-zhong Shang, Zi-rong Luo, Jin-wei Qiao, and Chen Cheng</i>	
Trench-Crossing Capability Analysis of a Reconfigurable Tracked Mobile Robot . . . . .	509
<i>Mo Hei, Jianzhong Shang, Zirong Luo, and Zhuo Wang</i>	
<b>Service Robotics</b>	
Automatic Cooking Robot with a Novel Feeding System . . . . .	519
<i>Weixin Yan, En Guang Guan, Wentao Ma, Zhuang Fu, and Yanzheng Zhao</i>	
Multi-sensor Based Autonomous Underwater Manipulator Grasp . . . . .	530
<i>Zhihu Xiao, Guohua Xu, Fuyuan Peng, Guo yuan Tan, Xiaolong Xu, Xiong Shen, and Bo Yang</i>	
The Targets Pursuit for Multi-robot System with Hybrid Wireless Sensor Networks . . . . .	538
<i>Kun Shi, Zhiqiang Cao, Wenwen Zhang, and Chao Zhou</i>	
A Fast Robot Path Planning Algorithm Based on Image Thinning . . . . .	548
<i>Ke Xu, Ruiqing Fu, Lei Deng, Yongsheng Ou, and Xinyu Wu</i>	
Automatic Micro-manipulation Based on Visual Servoing . . . . .	558
<i>Xiangjin Zeng, Yanduo Zhang, and Xinhan Huang</i>	
Self-calibration of a Stewart Parallel Robot with a Laserranger . . . . .	570
<i>Yu Liu, Hong Liu, Fenglei Ni, Wenfu Xu, and Feng Han</i>	
Real-Time Binaural Sound Source Localization Using Sparse Coding and SOM . . . . .	582
<i>Do-hyeong Hwang and Jong-suk Choi</i>	

Design and Control of a Hydraulic-Actuated Leg Exoskeleton for Load-Carrying Augmentation . . . . . 590  
*Heng Cao, Jun Zhu, Chunming Xia, Hong Zhou, Xiao Chen, and Yu Wang*

**Computer Vision and Applications**

Globally Optimal and Region Scalable CV Model on Automatic Target Segmentation . . . . . 600  
*Huijie Fan, Wei Dong, and Yandong Tang*

The Registration of UAV Down-Looking Aerial Images to Satellite Images with Image Entropy and Edges . . . . . 609  
*Baojie Fan, Yingkui Du, Linlin Zhu, and Yandong Tang*

Fast 3D Time of Flight Data Segmentation Using the U-V-Histogram Approach . . . . . 618  
*Thomas Schamm, Arne Rönnau, and J. Marius Zöllner*

Scale Invariant Feature Transform (SIFT) Parametric Optimization Using Taguchi Design of Experiments . . . . . 630  
*Dominic R. Maestas, Ron Lumia, Gregory Starr, and John Wood*

Robust Visual Monitoring of Machine Condition with Sparse Coding and Self-Organizing Map . . . . . 642  
*Haining Liu, Yanming Li, Nan Li, and Chengliang Liu*

An Evolving Machinery Fault Diagnosis Approach Based on Affinity Propagation Algorithm . . . . . 654  
*Nan Li, Yanming Li, Haining Liu, and Chengliang Liu*

Closed Form Solution for the Scale Ambiguity Problem in Monocular Visual Odometry . . . . . 665  
*Isaac Esteban, Leo Dorst, and Judith Dijk*

Recognizing 3D Human Motions Using Fuzzy Quantile Inference . . . . . 680  
*Mehdi Houry and Honghai Liu*

Evaluating the Fuzzy Coverage Model for 3D Multi-camera Network Applications . . . . . 692  
*Aaron Mavrinar, Jose Luis Alarcon Herrera, and Xiang Chen*

Emotion Recognition in Facial Image Sequences Using a Combination of AAM with FACS and DBN . . . . . 702  
*Kwang-Eun Ko and Kwee-Bo Sim*

## Sensor Design, Multi-sensor Data Fusion

An Analysis of Node Localization Error Behavior in Wireless Sensor Networks .....	713
<i>Dan Hai, Hui Zhang, and Zhiqiang Zheng</i>	
Consensus Target Tracking in Multi-robot Systems .....	724
<i>Zongyao Wang, Dongbing Gu, Tao Meng, and Yanzhi Zhao</i>	
A Decentralized Junction Tree Approach to Mobile Robots Cooperative Localization .....	736
<i>Hua Mu, Meiping Wu, Hongxu Ma, and Wenqi Wu</i>	
Optimized Particles for 3-D Tracking .....	749
<i>Huiying Chen and Youfu Li</i>	
Design of a Finger-Tip Flexible Tactile Sensor for an Anthropomorphic Robot Hand .....	762
<i>Yuan-Fei Zhang, Yi-Wei Liu, Ming-He Jin, and Hong Liu</i>	
<b>Author Index</b> .....	775

# Cartesian Impedance Control on Five-Finger Dexterous Robot Hand DLR-HIT II with Flexible Joint

Zhaopeng Chen<sup>1,2</sup>, Neal Y. Lii<sup>1</sup>, Minghe Jin<sup>2</sup>, Shaowei Fan<sup>2</sup>, and Hong Liu<sup>1,2</sup>

<sup>1</sup> Institute of Robotics and Mechatronics, German Aerospace Center, DLR, 82234 Wessling, Germany

<sup>2</sup> State Key Laboratory of Robotics and System, Harbin Institute of Technology, HIT, 150001 Harbin, China  
zhaopeng.chen@dlr.de

**Abstract.** This paper presents an impedance controller for five-finger dexterous robot hand DLR-HIT II, which is derived in Cartesian space. By considering flexibility in finger joints and strong mechanical couplings in differential gear-box, modeling and control of the robot hand are described in this paper. The model-based friction estimation and velocity observer are carried out with an extended Kalman filter, which is implemented with parameters estimated by Least Squares Method. The designed estimator demonstrates good prediction performance, as shown in the experimental results. Stability analysis of the proposed impedance controller is carried out and described in this paper. Impedance control experiments are conducted with the five-finger dexterous robot hand DLR-HIT II in Cartesian coordinates system to help study the effectiveness of the proposed controller with friction compensation and hardware architecture.

**Keywords:** dexterous robot hand, flexible joint, Cartesian space, impedance control, friction compensation.

## 1 Introduction

Since intelligence and hands are the two features that most distinguish humans from other animals, researchers have long been fascinated by the mechanical dexterity of the human hand. Nevertheless it is still an unmatched standard if a broad scope of manipulation tasks is considered [1]. It is highly anticipated that the forthcoming anthropomorphic robot hand will eventually supplant human labor in intricate and hazardous tasks execution.

Many multifingered robot hands have been developed mainly in two categories: With actuators mounted in the robot hand, or located outside the robot hand and power transmitted through tendon cables. The latter robot hands (e.g., Stanford-JPL hand by Salisbury et al. [2], the Utah/MIT hand by Jacobsen et al. [3], and the robonaut hand by Lovchik et al. [4]) suffer from the elasticity of the tendon cable with inaccurate joint angle control, and are problematic in

commercial manufacturing due to their mechanical complexity. However the former group of robot hands (e.g., the Gifu Hand by Kawasaki et al. [5], the DLR Hand II by Butterfass [6] and the DLR/HIT Hand I by Liu [7]) face other problems, such as insufficient fingers and joints, inappropriate large size and weight for practical task.

Manipulative dexterity and grasp robustness are two main types of functional requirements for a robot hand. By any reasonable definition, manipulation fundamentally requires mechanical interaction with the object being manipulated. The problems of controlling the mechanical interaction between a manipulator and its environment have been addressed by many researchers [8] [9] [10]. Hogan introduced a framework for impedance control [11], which can be used to achieve compliant manipulation and reliable grasping with different objects in unknown environments. Extensive literature exists on theories and application of impedance control law. Kazerooni et al. [12] proposed practical compliant impedance control law in Cartesian coordinates frame, by considering manipulator and environment as a whole system to preserve the global stability. A unified approach for motion and force control of robot manipulator in operational space with kinematic singularity treatment was designed by Khatib [13]. An internal force-based impedance control scheme for cooperating manipulators was introduced by Bonitz and Hisa [14]. Asymptotic stability was proven using Lyapunov theory and simulation results were presented to validate the proposed concepts.

The above approaches of impedance control were developed for robot systems with rigid body models, which neglect effects of joint elasticity. However, joint flexibility is non-neglectable in robots powered by DC or AC motor in series with harmonic drivers, which are common in many precision positioning and light weight applications. An effective Cartesian impedance control method based on singular perturbation model was proposed by Spong [15], in which the flexibility of joints is treated in a sufficiently fast inner torque control loop. Other decoupling-based approaches are given by Lin and Goldenberg [16], De Luca [17], which provide a linearized closed loop system, and ensure global asymptotic stability. However, these approaches resulted in moderate performance and robustness in joint torque sensor-based robots, due to the requirement on high derivative of link side positions and accurate robot model. To achieve both theoretically sound and practically feasible impedance controller for robots with elastic joints, Albu-schaeffer and Ott et al. [18] [19] introduced passivity-based impedance control which solely relies on motor position, joint torque, and their first-order derivatives. However, nonlinear effects such as friction compensation of the robots with elastic joint are not completely implemented in the above work. In this paper, a impedance control strategy for dexterous robot hand is proposed with friction compensation, which is estimated by extended Kalman filter.

This paper is organized as follows: Section 2 presents the dexterous robot hand system and robot model; Section 3 presents the joint and cartesian impedance controller of the dexterous robot hand, global asymptotic stability is also proven; Section 4 deals with the velocity observer and friction compensation based on

extended Kalman filter; experimental results are presented in Section 5. Finally, the conclusions of this work are drawn and presented in Section 6.

## 2 Robot Hand System and Dynamic Model

By considering the requirements for dexterous grasping and manipulation of objects, A five-fingered dexterous robot hand DLR-HIT II was developed jointly by German Aerospace Center, Germany and Harbin Institute of Technology, Chian [20], as illustrated in Fig. 1. The proposed hardware and software control architecture is composed of multisensory system, three DSP/FPGA based processing units for the distal controller, finger controller and palm controller, external real-time PC and multipoint real-time communication system.

To realize an anthropomorphic robot hand design, as well as augmentation of available torque for a given motor size and a reduced weight, the proximal joint of the robot finger is realized as coupled joints using a differential gear with a harmonic drive and timing belt. As shown in Fig. 2, the differential gear box is composed of four conical gears. Movement of one joint is realized by the coordinated movement of two actuators. Without considering the elasticity, the transformation matrix for each finger post differential gearing can be written as:

$$\theta_l = L\theta_q \quad (1)$$

$$\tau = L^T \tau_l \quad (2)$$

with

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix} \in \mathbb{R}^{3 \times 3} \quad (3)$$

$\theta_l$  and  $\theta_q$  denote the link side position and joint position expressed in joint coordinates, respectively.  $q$  represents  $\theta_q$  later in this article to indicate the

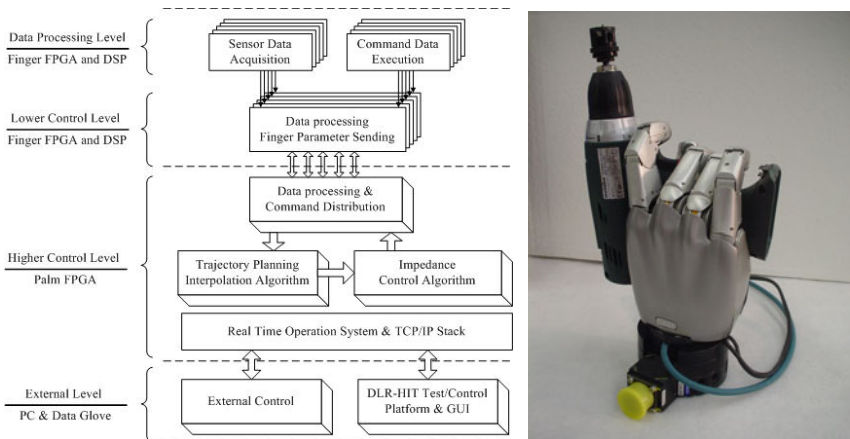
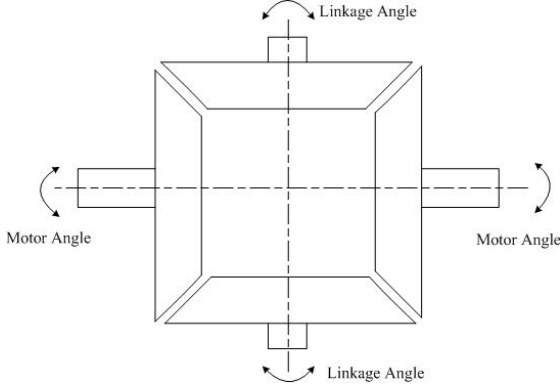


Fig. 1. Dexterous Robot Hand DLR-HIT II



**Fig. 2.** Modeling of Differential Bevel Gear

position of joint after link elasticity and joint coupling. Accordingly,  $\tau_l$  and  $\tau_q$  represent joint torques in joint and motor coordinates, while  $\tau_q$  will be denoted as  $\tau$  later in this paper.

Experimental results have shown that the elastic deformation of the joints, either through deformation of the bearing or of the gear teeth, would also introduce joint flexibility. Therefore the joint elasticity is taken into account in the modeling of robotic manipulators in this paper, which is modeled by a linear torsional spring with stiffness  $\mathbf{K}$ . The kinematics model of the robot hand with elastic joint and coupled joints can then be expressed as below [15]:

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{g}(\mathbf{q}) = \boldsymbol{\tau} + \boldsymbol{\tau}_{ext} \quad (4)$$

$$\mathbf{J}_m\ddot{\boldsymbol{\theta}}_m + \mathbf{L}^{-T}\boldsymbol{\tau} + \boldsymbol{\tau}_{f,m} = \boldsymbol{\tau}_m \quad (5)$$

$$\boldsymbol{\tau} = \mathbf{K}(\mathbf{L}\mathbf{q} - \boldsymbol{\theta}_m) \quad (6)$$

where  $\mathbf{M}(\mathbf{q})$ ,  $\mathbf{C}(\mathbf{q})$  and  $\mathbf{g}(\mathbf{q})$  represent the inertia matrices, centrifugal term, and gravity term, respectively. The joint torque vector is given by  $\mathbf{K}(\mathbf{L}\mathbf{q} - \boldsymbol{\theta}_m)$ , where  $\boldsymbol{\theta}_m$  indicates the vector of the motor angle divided by the gear ratio, and  $\mathbf{q}$  represents the link side joint angle.  $\mathbf{K}$ ,  $\mathbf{J}_m$  are diagonal matrices which contain the joint stiffness, and the motor inertia multiplied by the gear ratio squared.  $\boldsymbol{\tau}_{ext}$  and  $\boldsymbol{\tau}_{f,m}$  are external torque vector and friction torque vector, respectively. The generalized actuator torque vector,  $\boldsymbol{\tau}_m$ , is considered as the control input.

### 3 Impedance Controller Design for Dexterous Robot Hand

The goal of the impedance controller is to achieve a desired dynamic behavior with respect to external forces and torques acting on the link side of the end



effector. As described in last section, robot hand dynamics can be expressed in the joint coordinates system by using a transformation matrix  $\mathbf{L}$ . Since elasticity in the joints of DLR-HIT II robot hand is not negligible, resulting dynamics of a motor position-based PD controller is influenced by the joint elasticity and the motor inertia. By a negative feedback of the joint torque  $\boldsymbol{\tau}$ , the apparent inertia can be scaled down such that the closed-loop system reacts to external forces  $\mathbf{F}_{ext}$  as if the robot inertia were smaller [18], which can be represented as:

$$\boldsymbol{\tau}_j = \mathbf{J}\mathbf{J}_c^{-1}\mathbf{u} + (\mathbf{I} - \mathbf{J}\mathbf{J}_c^{-1})\boldsymbol{\tau} \quad (7)$$

with

$$\boldsymbol{\tau}_j = \mathbf{L}^T(\boldsymbol{\tau}_m - \boldsymbol{\tau}_{f,m}) \quad (8)$$

$$\mathbf{J} = \mathbf{L}^T\mathbf{J}_m\mathbf{L} \quad (9)$$

where  $\mathbf{u} \in \mathbb{R}$  is the new input variable, and  $\mathbf{J}_c$  is a diagonal, positive definite matrix with  $j_{ci} < j_i$ .

The position and orientation of the end-effector can be described by a set of local coordinates  $\mathbf{x} \in \mathbb{R}^m$ , and the relationship between Cartesian coordinates  $\mathbf{x}$  and the configuration coordinates  $\mathbf{q} \in \mathcal{Q}$  is given by a known function  $\mathbf{f} : \mathcal{Q} \rightarrow \mathbb{R}^m$ , i.e.  $\mathbf{x} = \mathbf{f}(\mathbf{q})$ . With the Jacobian  $\mathbf{J}(\mathbf{q}) = \partial\mathbf{f}(\mathbf{q})/\partial\mathbf{q}$ , Cartesian velocities and accelerations can be expressed as

$$\dot{\mathbf{x}} = \mathbf{J}(\mathbf{q})\dot{\mathbf{q}} \quad (10)$$

$$\ddot{\mathbf{x}} = \mathbf{J}(\mathbf{q})\ddot{\mathbf{q}} + \dot{\mathbf{J}}(\mathbf{q})\dot{\mathbf{q}} \quad (11)$$

Throughout this paper only the nonsingular case is considered, thus it is assumed that the manipulator's Jacobian  $\mathbf{J}(\mathbf{q})$  has full row rank in the considered region of the workspace.

To specify the desired impedance behavior, the position error  $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{x}_d$ , between real position  $\mathbf{x}$  and a virtual equilibrium position (possibly time-varying)  $\mathbf{x}_d$ , is introduced. The goal of the impedance controller design here is to alter the system dynamics (4) such that, in the presence of external forces and torques at the end-effector  $\mathbf{F}_{ext} \in \mathbb{R}^m$ , a dynamic relationship between  $\tilde{\mathbf{x}}$  and  $\mathbf{F}_{ext}$  could be achieved as follows:

$$\Lambda_d\ddot{\tilde{\mathbf{x}}} + D_d\dot{\tilde{\mathbf{x}}} + K_d\tilde{\mathbf{x}} = \mathbf{F}_{ext} \quad (12)$$

where  $\Lambda_d$ ,  $D_d$  and  $K_d$  are the symmetric and positive definite matrices of the desired inertia, damping and stiffness, respectively.

The relationship between the external torque vector  $\boldsymbol{\tau}_{ext}$  and the generalized external force vector  $\mathbf{F}_{ext}$  on the end-effector is given by:

$$\boldsymbol{\tau}_{ext} = \mathbf{J}(\mathbf{q})^T\mathbf{F}_{ext} \quad (13)$$

By substituting (10), (11) and (13) into (4) and rewriting  $\mathbf{g}(\mathbf{q})$ ,  $\boldsymbol{\tau}$  in the form of the equivalent task space as  $\mathbf{F}_g(\mathbf{x})$ ,  $\mathbf{F}_\tau$ , the relationship between the Cartesian coordinates  $\mathbf{x}$  and the joint torques  $\boldsymbol{\tau}$  can be expressed as:

$$\Lambda(\mathbf{x})\ddot{\mathbf{x}} + \boldsymbol{\mu}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}} + \mathbf{F}_g(\mathbf{x}) = \mathbf{F}_\tau + \mathbf{F}_{ext} \quad (14)$$

where  $\Lambda(\mathbf{x})$  and  $\boldsymbol{\mu}(\mathbf{x}, \dot{\mathbf{x}})$  are the inertia matrix and the Coriolis/centrifugal matrix with respect to the coordinates  $\mathbf{x}$ , and given by:

$$\Lambda(\mathbf{x}) = \mathbf{J}(\mathbf{q})^{-T} \mathbf{M}(\mathbf{q}) \mathbf{J}(\mathbf{q})^{-1} \quad (15)$$

$$\boldsymbol{\mu}(\mathbf{x}, \dot{\mathbf{x}}) = \mathbf{J}(\mathbf{q})^{-T} (\mathbf{C}(\mathbf{q}, \dot{\mathbf{q}}) - \mathbf{M}(\mathbf{q}) \mathbf{J}(\mathbf{q})^{-1} \dot{\mathbf{J}}(\mathbf{q})) \mathbf{J}(\mathbf{q})^{-1} \quad (16)$$

with  $\mathbf{q} = \mathbf{f}^{-1}(\mathbf{x})$  and  $\dot{\mathbf{q}} = \mathbf{J}(\mathbf{f}^{-1}(\mathbf{x}))\dot{\mathbf{x}}$ .

Combining (14) and (12), the impedance control law, which is the desired closed loop system, with  $\mathbf{F}_\tau$  as the control input, can be arrived:

$$\mathbf{F}_\tau = \Lambda(\mathbf{x})\ddot{\mathbf{x}}_d + \boldsymbol{\mu}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}} + (\Lambda(\mathbf{x})\Lambda_d^{-1} - \mathbf{I})\mathbf{F}_{ext} \quad (17)$$

$$+ \mathbf{F}_g(\mathbf{x}) - \Lambda(\mathbf{x})\Lambda_d^{-1}(\mathbf{D}_d\dot{\tilde{\mathbf{x}}} + \mathbf{K}_d\tilde{\mathbf{x}}) \quad (18)$$

If the desired torque vector  $\boldsymbol{\tau}$  is chosen as:

$$\boldsymbol{\tau} = \mathbf{J}(\mathbf{q})^T \mathbf{F}_\tau + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} - \mathbf{J}(\mathbf{q})^T \Lambda(\mathbf{x}) \dot{\mathbf{J}}(\mathbf{q}) \mathbf{J}(\mathbf{q})^{-1} \quad (19)$$

With the assumption that centripetal and Coriolis forces can be ignored at the robot's relatively low operating speeds. Furthermore, If the desired inertia  $\Lambda_d$  is chosen as identical to the robot inertia  $\Lambda(\mathbf{x})$ , the feedback of external forces  $\mathbf{F}_{ext}$  can be avoided. This results in the following actual implementation of the impedance controller:

$$\mathbf{F}_\tau = \Lambda(\mathbf{x})\ddot{\mathbf{x}}_d - \mathbf{D}_d\dot{\tilde{\mathbf{x}}} - \mathbf{K}_d\tilde{\mathbf{x}} + \mathbf{F}_g(\mathbf{x}) \quad (20)$$

and the desired joint torques  $\boldsymbol{\tau}$ :

$$\boldsymbol{\tau} = \mathbf{g}(\mathbf{q}) + \mathbf{J}(\mathbf{q})^T (\Lambda(\mathbf{x})\ddot{\mathbf{x}}_d - \mathbf{D}_d\dot{\tilde{\mathbf{x}}} - \mathbf{K}_d\tilde{\mathbf{x}}) \quad (21)$$

Using motor  $\boldsymbol{\theta}$  instead of the link side angles  $\mathbf{q}$  in the forward kinematics  $\mathbf{x} = \mathbf{f}(\mathbf{q})$ , impedance controller based on PD position control in the Cartesian coordinates can be generalized. The feedback law is given by:

$$\begin{aligned} \mathbf{u} &= -\mathbf{J}(\boldsymbol{\theta})^T (\mathbf{K}_x \tilde{\mathbf{x}}(\boldsymbol{\theta}) + \mathbf{D}_x \dot{\tilde{\mathbf{x}}}) \\ \tilde{\mathbf{x}}(\boldsymbol{\theta}) &= \mathbf{f}(\boldsymbol{\theta}) - \mathbf{x}_d \\ \dot{\tilde{\mathbf{x}}} &= \mathbf{J}(\boldsymbol{\theta})\dot{\boldsymbol{\theta}} \end{aligned} \quad (22)$$

with  $\mathbf{K}_x$  and  $\mathbf{D}_x$  represent the desired stiffness and damping matrices, respectively corresponding to  $\mathbf{K}_d$  and  $\mathbf{D}_d$  in (21).  $\mathbf{x}_d$  indicates the virtual motor side position in Cartesian coordinates. As such, the controller in (22), together with (7), forms the closed loop system:

$$\begin{aligned} \mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}}) + \mathbf{g}(\mathbf{q}) &= \boldsymbol{\tau} + \boldsymbol{\tau}_{ext} \\ \mathbf{J}_c \ddot{\boldsymbol{\theta}} + \mathbf{J}(\boldsymbol{\theta})^T (\mathbf{K}_x \tilde{\mathbf{x}}(\boldsymbol{\theta}) + \mathbf{D}_x \dot{\tilde{\mathbf{x}}}) + \boldsymbol{\tau} &= \bar{\mathbf{g}}(\boldsymbol{\theta}) \end{aligned} \quad (23)$$

Since link side deviations from the steady state positions due to gravity is not negligible, the following Lyapunov function  $V(\mathbf{q}, \dot{\mathbf{q}}, \boldsymbol{\theta}, \dot{\boldsymbol{\theta}})$  is chosen:

$$\begin{aligned} V(\mathbf{q}, \dot{\mathbf{q}}, \boldsymbol{\theta}, \dot{\boldsymbol{\theta}}) &= \frac{1}{2} \dot{\mathbf{q}}^T \mathbf{M}(\mathbf{q}) \dot{\mathbf{q}} + \frac{1}{2} \dot{\boldsymbol{\theta}}^T \mathbf{J}_c \dot{\boldsymbol{\theta}} + \frac{1}{2} (\boldsymbol{\theta} - \mathbf{q})^T \mathbf{K} (\boldsymbol{\theta} - \mathbf{q}) \\ &\quad + \frac{1}{2} \tilde{\mathbf{x}}(\boldsymbol{\theta})^T \mathbf{K}_x \tilde{\mathbf{x}}(\boldsymbol{\theta}) + V_g(\mathbf{q}) - V_{\bar{g}}(\boldsymbol{\theta}) \\ &\geq 0 \end{aligned} \quad (24)$$

For  $\tau_{ext} = \mathbf{0}$ , derivative of Lyapunov function can be manipulated:

$$\mathbf{V}(q, \dot{q}, \theta, \dot{\theta}) = -\dot{x}^T D_x \dot{x} \leq 0 \quad (25)$$

The system converges to invariant set contained in subspace  $\mathbf{P} = (\theta_s, \mathbf{0}, q_s, \mathbf{0})^T$ , which is actually the steady state given by (23):

$$\mathbf{K}(\theta_s - q_s) = g(q_s) \quad (26)$$

$$\mathbf{K}(\theta_s - q_s) + \mathbf{J}(\theta_s)^T \mathbf{K}_x \tilde{x}(\theta_s) = \bar{g}(\theta) \quad (27)$$

Then the asymptotic stability is shown by the LaSalle Invariance Principle.

## 4 Velocity Observer and Friction Compensation

### 4.1 Friction Model Linearization and Parameters Estimation

Performance of the proposed impedance controller can be hindered by nonlinear friction. In this paper, friction model  $\tau_{f,m}$  is composed of Coulomb friction  $\tau_c$  and viscous friction  $\tau_v$ , which can be written as:

$$\tau_{f,m} = \tau_c + \tau_v = \alpha \text{sign}(\dot{\theta}_m) + \beta \dot{\theta}_m \quad (28)$$

with  $\alpha$  and  $\beta$  denoting the Coulomb and viscous coefficient, respectively. The following differential equation represents the resulting dynamics of the system with the friction model:

$$\mathbf{J}_c \ddot{\theta} + \alpha \text{sign}(\dot{\theta}_m) + \beta \dot{\theta}_m = \tau_m \quad (29)$$

In order to implement a controller with friction compensation, it is necessary to determine parameters corresponding to the robot dynamic model. Nonlinearity of the above friction model resides in the signum function term  $\text{sign}(\dot{\theta})$ . By multiplying both sides of the equation (29) with term  $\dot{\theta}$ , the friction model is given by linear form:

$$\mathbf{J}_c \dot{\theta} \ddot{\theta} + \beta \dot{\theta}^2 + \underbrace{\alpha \dot{\theta} \text{sign}(\dot{\theta})}_{|\dot{\theta}|} = \dot{\theta} \tau \quad (30)$$

By integrating (30) over one timestep  $t_s = t_{k+1} - t_k$  with a trapezoidal approximation, the least-squares estimation can be achieved:

$$\underbrace{\begin{bmatrix} \frac{1}{2}(\dot{\theta}_{k+1}^2 - \dot{\theta}_k^2) & \frac{t_s}{2}(\dot{\theta}_{k+1}^2 + \dot{\theta}_k^2) & \frac{t_s}{2}(|\dot{\theta}_{k+1}^2| + |\dot{\theta}_k^2|) \\ \downarrow k=0 \dots N \\ \vdots \end{bmatrix}}_{LS} = \begin{bmatrix} \mathbf{J}_c \\ \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} \frac{t_s}{2}(\tau_{k+1} \dot{\theta}_{k+1} - \tau_k \dot{\theta}_k) \\ \downarrow k=0 \dots N \\ \vdots \end{bmatrix}$$

## 4.2 Velocity Observer and Friction Estimation

The extended Kalman filter (EFK) is adopted in this paper to estimate velocity and friction, which is subdivided into prediction step and estimation step as shown in [21]. Based on the current state and the dynamic model of the system, a forecast can be calculated for the state in the prediction step:

$$\begin{aligned}\Delta\hat{x}_{k+1}^- &= A_k\Delta\hat{x}_k^+ + B_K u_k \\ P_{k+1}^- &= A_k P_k^+ + G_K Q_k G_K^T\end{aligned}\quad (31)$$

The estimation step is defined, where the forecast and the measurements are compared, to arrive the following:

$$\begin{aligned}K_k &= P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \\ \Delta\hat{x}_k^+ &= \Delta\hat{x}_k^- - K (H_k \Delta\hat{x}_k^- - \Delta y_k) \\ P_k^+ &= (I - K_k H_k) P_k^-\end{aligned}\quad (32)$$

In order to eliminate a possible divergence of  $\Delta x$ , the error state is set to zero after the estimation step.

To derive an extended Kalman filter, viscous and static friction parameters  $\mathbf{b}$  and  $\mathbf{c}$  are modeled as constant system states.  $\tau$  is considered as a state variable, or specifically a measurement variable rather than an input variable. The dynamic model of the system can be expressed as [22]:

$$\underbrace{\frac{d}{dt} \begin{pmatrix} \theta \\ \dot{\theta} \\ \tau \\ \mathbf{b} \\ \mathbf{c} \end{pmatrix}}_{:=x} = \underbrace{\begin{pmatrix} \dot{\theta} \\ \frac{1}{J_C}(\tau - \mathbf{b}\dot{\theta} - \mathbf{c}\text{sign}(\dot{\theta})) \\ 0 \\ 0 \\ 0 \end{pmatrix}}_{:=a}\quad (33)$$

By partial derivation of the system dynamics equation and replacing the partial derivation sign  $\partial$  with differences sign  $\Delta$ . The following linearized system dynamics equation can be achieved:

$$\frac{d}{dt} \begin{pmatrix} \Delta\theta \\ \Delta\dot{\theta} \\ \Delta\tau \\ \Delta\mathbf{b} \\ \Delta\mathbf{c} \end{pmatrix} = A_{lin} \cdot \begin{pmatrix} \Delta\theta \\ \Delta\dot{\theta} \\ \Delta\tau \\ \Delta\mathbf{b} \\ \Delta\mathbf{c} \end{pmatrix}\quad (34)$$

where:

$$A_{lin} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & -(\frac{b}{J_c} + \alpha \frac{f_c}{J_c}) & \frac{1}{J_c} & \frac{-\dot{\theta}}{J_c} & \frac{-f(\dot{\theta})}{J_c} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (35)$$

$$f(\dot{\theta}) = \begin{cases} -1 & \dot{\theta} < -limit \\ \frac{1}{limit} & -limit < \dot{\theta} < limit \\ 1 & limit < \dot{\theta} \end{cases} \quad (36)$$

with  $\alpha$  represents the derivation of  $f(\dot{\theta})$ , given by:

$$\alpha = \begin{cases} 0 & \dot{\theta} < -limit \\ \frac{1}{limit} & -limit < \dot{\theta} < limit \\ 0 & limit < \dot{\theta} \end{cases}$$

Considering that  $\alpha$  and  $\beta$  are constant and estimated with least squares in the prior section, the linearized extended Kalman filter can be achieved by solving (34):

$$\Delta x_{k+1} = \begin{pmatrix} 1 & a_2 & a_3 \\ 0 & a_1 & \frac{1}{J_c} \cdot a_2 \\ 0 & 0 & 1 \end{pmatrix} \cdot \Delta x_k \quad (37)$$

Where:

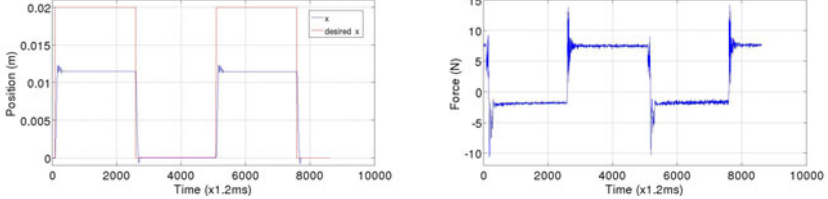
$$\begin{aligned} a_1 &= e^{-(b+\alpha c) \cdot \frac{t_s}{J_c}} \\ a_2 &= \frac{J_c}{b + \alpha c} \cdot (1 - e^{-(b+\alpha c) \cdot \frac{t_s}{J_c}}) \\ a_3 &= \frac{J_c}{(b + \alpha c)^2} \cdot (e^{-(b+\alpha c) \cdot \frac{t_s}{J_c}} - 1 + (b + \alpha c) \frac{t_s}{J_c}) \end{aligned}$$

## 5 Experiments

The Cartesian impedance control experiment is carried out in the five-finger dexterous robot hand DLR-HIT II with a hard real time control cycle 200  $\mu s$ . Multi-point LVDS serial communication system and QNX real time OS are utilized to fulfill the requirements of the proposed controller.  $D_x$  and  $K_k$  are designed by the double-diagonalization approach with the robot inertia matrix and the desired damping ratio, as presented in [23]. Other parameters needed for implementing the impedance controller (23) can be generated by directly using Pro/E model of the dexterous hand.

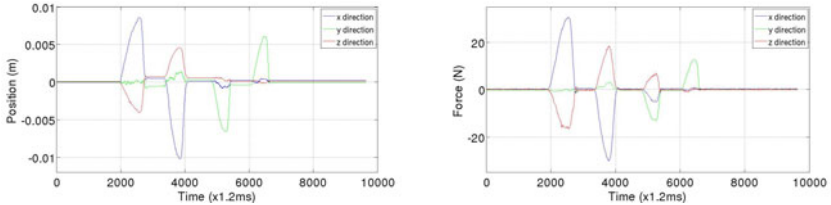
In the following two experiments, the former one is conducted to test the performance of the controller, whereas the latter one is carried out to show the compliant behavior of the robot hand. The designed Cartesian impedance controller is implemented in a single finger of the robot hand. The finger tracks

the desired position trajectory (red line) and makes contact with a rigid external object at the position offset  $\Delta x = 0.011m$  in the  $x$  direction, as shown in Fig. 3. Together with Cartesian force response in Fig. 3, the experimental results show that the proposed impedance controller is effective in position tracking.



**Fig. 3.** Position tracking and force response

As shown in Fig. 4, The robot overcomes the gravity and friction, returning to the equilibrium Cartesian position  $x_d$  as soon as the external force is released. With the friction and gravity compensation proposed in this paper, the static error in the  $x$  axis is less than 0.2 mm, as well as  $y$  and  $z$  direction. It can therefore be concluded that the Cartesian impedance controller is successfully realized.



**Fig. 4.** Cartesian coordinate position and force response with impedance controller

## 6 Conclusions

In this paper, a Cartesian impedance controller with friction compensation is derived for the dexterous robot hand DLR-HIT II hand with flexible joints. Global asymptotic stability is guaranteed by LaSalle Principle analysis. To improve the performance of the proposed controller, model based friction compensation is adopted in this paper, in which friction parameters are estimated with the Least Squares Method. Together with joint velocity observed by extended Kalman filter, non-linear friction compensation can be derived. Two experiments are carried out on the DLR-HIT hand to show the effectiveness of designed impedance controller with friction compensation and its compliant behavior with the robot hand. For the harmonic drive robot hand with joint torque feedback, accurate position tracking and stable torque/force response can be achieved with the proposed Cartesian impedance controller.

## References

1. Bicchi, A.: Hands for dexterous manipulation and robust grasping: a difficult road toward simplicity. *IEEE Transactions on Robotics and Automation* 16(6), 652–662 (2000)
2. Salisbury, J., Craig, J.: Articulated hands: Force control and kinematic issues. *The International Journal of Robotics Research* 1(1), 4 (1982)
3. Jacoben, S., et al.: The Utah/MIT dexterous hand: Work in progress. *Int. J. Robot* 3(4), 21–50 (1984)
4. Lovchik, C., Difler, M.: The robonaut hand: A dextrous robotic hand for space. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 907–912 (1999)
5. Kawasaki, H., Komatsu, T., Uchiyama, K.: Dexterous anthropomorphic robot hand with distributed tactile sensor: Gifu hand II, IEEE. *ASME Trans. on Mechatronics* 7(3), 296–303 (2002)
6. Butterfass, J., Grebenstein, M., Liu, H.: DLR-Hand II: Next generation of a dexterous robot hand. In: *Proceedings of the 2001 IEEE International Conference on Robotics & Automation*, pp. 109–114 (2001)
7. Liu, H., Meusel, P., Hirzinger, G., Jin, M., Liu, Y., Xie, Z.: The modular multisensory DLR-HIT-hand: Hardware and software architecture. *IEEE/ASME Transactions on Mechatronics* 13(4), 461–469 (2008)
8. Chiaverini, S., Sciavicco, L.: The parallel approach to force/position control of robotic manipulators. *IEEE Transactions on Robotics and Automation* 9(4), 361–373 (1993)
9. Volpe, R., Khosla, P.: A theoretical and experimental investigation of explicit force control strategies for manipulators. *IEEE Transactions on Automatic Control* 38(11), 1634–1650 (1993)
10. Roy, J., Whitcomb, L.: Adaptive force control of position/velocity controlled robots: theory and experiment. *IEEE Transactions on Robotics and Automation* 18(2), 121–137 (2002)
11. Hogan, N.: Impedance control—An approach to manipulation. I—Theory. II—Implementation. III—Applications. *ASME, Transactions, Journal of Dynamic Systems, Measurement and Control* 107 (1985), ISSN: 0022-0434
12. Kazerooni, H., Houpt, P., Sheridan, T.: Robust compliant motion for manipulators: the fundamental concepts of compliant motion (part I); design method (part II). *IEEE Journal of Robotics and Automation (legacy, pre-1988)* 2(2), 83–105 (1986)
13. Khatib, O.: A unified approach for motion and force control of robot manipulators: The operational space formulation. *IEEE Journal of Robotics and Automation* 3(1), 43–53 (1987)
14. Bonitz, R., Hsia, T.: Internal force-based impedance control for cooperating manipulators. *IEEE Transactions on Robotics and Automation* 12(1), 78–89 (1996)
15. Spong, M., Khorasani, K., Kokotovic, P.: An integral manifold approach to the feedback control of flexible joint robots. *IEEE Journal of Robotics and Automation (legacy, pre-1988)* 3(4), 291–300 (1987)
16. Tian, L., Goldenberg, A.: Robust adaptive control of flexible joint robots with joint torque feedback. In: *Proceedings of 1995 IEEE International Conference on Robotics and Automation*, vol. 1 (1995)
17. De Luca, A., Lucibello, P.: A general algorithm for dynamic feedback linearization of robots with elastic joints. In: *IEEE International Conference on Robotics and Automation*, pp. 504–510 (1998)

18. Ott, C., Albu-Schaeffer, A., Kugi, A., Hirzinger, G.: On the Passivity-Based Impedance Control of Flexible Joint Robots. *IEEE Transactions on Robotics* 24(2), 416–429 (2008)
19. Albu-Schaeffer, A., Ott, C., Hirzinger, G.: A unified passivity-based control framework for position, torque and impedance control of flexible joint robots. *The International Journal of Robotics Research* 26(1), 5–21 (2007)
20. Liu, H., Wu, K., Meusel, P., Seitz, N., Hirzinger, G., Jin, M., Liu, Y., Fan, S., Lan, T., Chen, Z.: Multisensory five-finger dexterous hand: The DLR/HIT Hand II. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2008*, pp. 3692–3697 (2008)
21. Gelb, A.: *Applied optimal estimation*. MIT Press, Cambridge (2002)
22. Connette, C.P.: *Intern Report On DLR Hand II*, Institute of Robotic and Mechatronics, DLR (2006)
23. Albu-Schaeffer, A., Ott, C., Hirzinger, G.: A passivity based Cartesian impedance controller for flexible joint robots—Part II: full state feedback, impedance design and experiments. In: *IEEE International Conference on Robotics and Automation, Citeseer*, pp. 2666–2672 (2004)



# Changeable Grasping Force Unit for Under-Actuated Gripper with Self-adaptive Grasping

Deyang Zhao and Wenzeng Zhang

Key Laboratory for Advanced Materials Processing Technology, Ministry of Education;  
Dept. of Mechanical Engineering, Tsinghua University, Beijing 100084, China  
zdy07@mails.tsinghua.edu.cn

**Abstract.** The grasping force of traditional under-actuated hands can hardly be kept during grasping. It is a crucial task to improve the grasping forces of robotic hands. Changeable grasping forces unit is one of the most effective solutions. This paper proposed a linkage changeable grasping force (CGF) unit integrated with linkage under-actuation mechanism in accordance with summary of proceeding work of CGF units in TH series. The inner space of the finger is fully used by the CGF unit. Analysis is conducted to explain relation between changeable forces and corresponding parameters. As a consequence, the CGF unit presented in this paper not only provides changeable forces for TH-3L Hand but also throw light on design approach which combines CGF unit and under-actuation.

**Keywords:** bionic dexterous hand, robot mechanism design, under-actuated grasping, self-adaptive grasping, changeable grasping force mechanism.

## 1 Introduction

Under-actuated hand was proposed as simplification of dexterous hands with fewer actuators and simpler control. It refers to a hand in which the number of actuators is less than the number of DOFs. The principle of under-actuation was presented in [1, 2]. It was pointed out that under-actuation in hands leads to self-adaptability. Most design are based on linkage or tendon cable, several of which are SSL Hand<sup>[3]</sup>, DIES-DIEM Hand<sup>[4]</sup>, Cassino Hand<sup>[5]</sup>, LARM Hand<sup>[6]</sup> and TH-3R Hand<sup>[7]</sup>.

A crucial problem in robotic hands design is how to keep the grasping force after the hand has grasped the object. The grasping force need to be kept to ensure firm grasp. The question is whether the force could be remained while the actuator stops outputting torque. To solve this problem, one method is to design robotic hands with changeable grasping force (CGF) unit. Several designs haven adopted electric elements to achieve changeable forces. In [8], each finger in the hand is actuated by tendon. At the end of each tendon, a spring connects the tendon to the palm of the force. As a result, the forces exerted by the fingers are changeable to the extension of the springs. In [9], the hand is also tendon based. However, there is a spring not connected to the tendon directly but to the wheel that tensions the tendon. In [10] which also a tendon hand presented, a part of the tendon is spring. All these design realize the changeable grasping forces. The disadvantages lie in all the designs are the forces are limited and the tendons

give rise to friction and compliance. Robotic hands with changeable forces not based on tendon cable can be found in [11] and [12]. In [11], each finger is connected by a spring to a block fixed in the palm. One function of the springs is to provide five fingers with changeable forces. The index finger of TH-1 hand except for the thumb is driven by a leaf spring in [12].

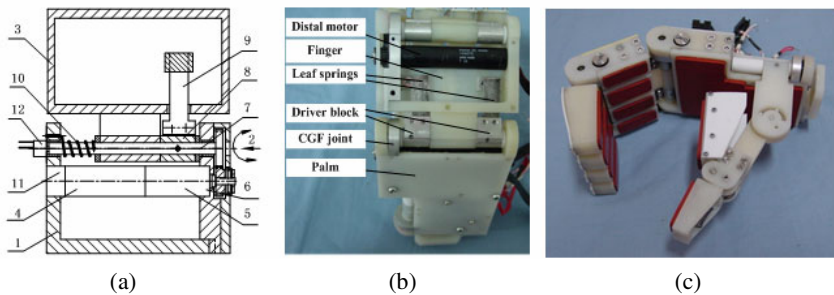
In this paper, CGF unit design in four TH series robot hands is addressed. A novel CGF unit combined with under-actuation mechanism is designed based on linkage for TH-3L Hand.

## 2 The Changeable Grasping Force Units in TH Series Hands

Applications of CGF unit include novel design of integral CGF unit or improvement of existing robotic hands. The former is more flexible since the hand design is able adapt to the CGF unit. The latter is relatively limited because the inter space of finger has been determined. TH-1 Hand, TH-2 Hand, TH-GCUA Hand and TH-COSA Hand are all under-actuated robotic hands designed by Tsinghua University. TH-1 Hand is based on changeable grasping forces in design. A universal CGF unit design approach was put forward in TH-2 Hand design, adopted in TH-GCUA Hand and TH-COSA Hand to achieve superior grasping forces.

### 2.1 The CGF Unit in TH-1 Hand

TH-1 Hand is the first hand designed with CGF unit. It begins CGF unit design in TH-series hands and forms foundation for latter designs. It has two fingers, one index with under-actuation and one index with CGF unit. The structure of the index is shown in Fig. 1.



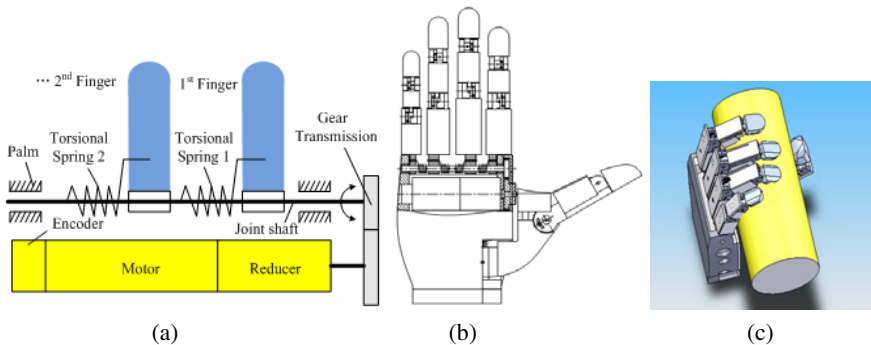
**Fig. 1.** Structure of the index finger in TH-1 Hand. 1-palm; 2-CGF joint; 3-finger; 4-motor; 5-reducer; 6-gear; 7-gear shaft; 8-driver block; 9-leaf spring; 10-torsional spring; 11-encoder; 12-potentiometer.

The index is composed of base segment, CGF Joint and CGF Segment. CGF Segment includes motor, joint shaft, driver blocks and leaf springs. The motor drives the driver blocks to bend through gear transmission mechanism. The leaf springs actuate the CGF Segment to rotate until it touches the surface of object. When CGF segment stops, the CGF segment has already rotated by an angle. The motor rotates continuously, and then

the joint shaft and driver blocks rotate continuously, which makes the leaf springs bend because CGF segment is restrained by the object. Hence elastic forces i.e. changeable forces are produced by the leaf springs. The motor stops when the elastic forces are strong enough. The joint will not rotate back due to the self-locked feature of the transmission mechanism. The distortion of leaf springs and elastic forces are kept, and the grasping forces are maintained.

## 2.2 The CGF Unit in TH-2 Hand

TH-2 Hand is developed in accordance with TH-1 Hand. The hand uses only two motors actuating five relative independent fingers, four of which are designed with CGF units. A universal approach of CGF units is adopted in TH-2 Hand for the first time shown in Fig. 2.



**Fig. 2.** Design of CGF unit in TH-2 Hand. (a)CGF Design; (b)CGF joint in TH-2 Hand; (c)Model of TH-2 Hand.

The torsional springs are set around root joint shaft, with one end fixed into the root joint shaft and the other end fixed into a corresponding finger. The springs make the four fingers that driven by a single motor independent to each other relatively. Fingers in Fig. 2 could be any type no matter under-actuated or dexterous so that this design can be adapted to different types of robotic hands. While robot hand grasps an object, some fingers might touch the object firstly during the rotation of root joint shaft. After these fingers are restrained, the torsional springs connected to them continue rotating thus deform by an angle. Grasping forces are maintained due to the deformation. Other fingers continue to rotate until all fingers grasp the object.

## 2.3 The CGF Units in TH-GCUA Hand and TH-COSA Hand

TH-GCUA Hand is based on the gesture-changeable under-actuated function. All five fingers are added CGF units that are the same with those in TH-2 Hand. Each finger has a torsional spring around the proximal joint shaft linking the active pulley and the transmission wheel. The proximal joint shaft drives the finger through the transmission wheel, torsional spring and active pulley.

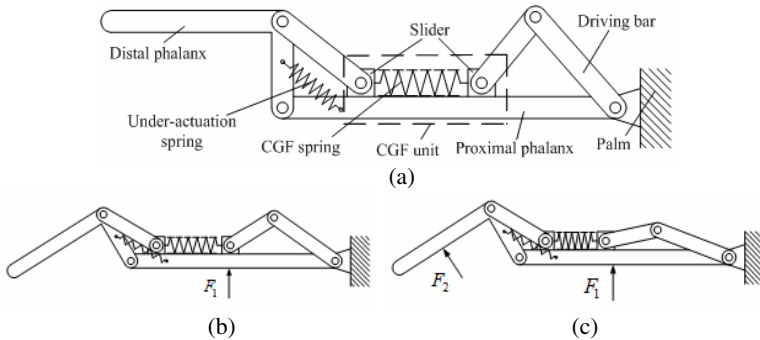
TH-COSA Hand is designed based on concept of coupling and self-adaptation (COSA). Linkage mechanism is applied to achieve COSA. CGF unit is added to each

finger also in the same approach with TH-2 Hand. One end of the torsional spring is fixed in the proximal joint shaft while the other end is fixed to the driving bar that actuates the finger.

### 3 Design of a Novel CGF Unit in TH-3L Hand

#### 3.1 Design of Linkage CGF Unit

Linkage mechanism is used to realize under-actuation in TH-3L Hand design. It is designed that the CGF unit is combined with under-actuation linkage. The CGF unit consists of a spring and two sliders. A compression spring will generate and keep changeable grasping forces. Sliders are used to convert rotation to parallel moving, avoiding interference between CGF unit and under-actuation mechanism. The structure of a two-DOF finger with CGF unit is shown in Fig. 3.



**Fig. 3.** Design of CGF unit in the TH-3L finger

The finger is composed of linkage mechanism, two sliders and two springs. One spring is designed for under-actuation, called under-actuation spring, and the other is CGF spring. The two terminals of under-actuation spring are fixed in the proximal phalanx and the distal phalanx, respectively. The CGF spring connects the lower slider to the upper slider. When the finger grasps an object, it keeps straight and rotates as a rigid body at first shown in Fig. 3(a). Then the under-actuation spring begins to extend after the proximal phalanx is blocked by the object, which means the distal phalanx rotates with respect to the proximal phalanx shown in Fig. 3(b). The distal phalanx continues rotating until it is constrained by the object either. The motor goes on working to compress the CGF spring here. The lower slider is pushed forward while the upper slider stays still. The forces produced by the CGF spring are the changeable grasping forces. The motor can be stopped when the CGF spring is compressed to the extreme shown in Fig. 3(c). Due to the self-locking of reducer, the whole mechanism will not move. Grasping forces are maintained as a result.

#### 3.2 Comparing of TH-3L Hand with Former TH Series Hands

In this part, design of TH-3L Hand will be compared with previous TH series hands. A summary will be given according to the comparison. Table 1 provides certain

**Table 1.** TH-series robotic hands

Hand	TH-1	TH-2	TH-GCUA	TH-COSA	TH-3L
<b>Fingers</b>	2	5	5	5	5
<b>Fingers with CGF</b>	1	4	5	5	5
<b>Under-actuation mechanism</b>	Active board and gears	Slip blocks	Belts and pulleys	Linkages	Linkages
<b>CGF unit</b>	Leaf springs and driver block	Torsional springs	Torsional springs	Torsional springs	Springs and sliders
<b>Relation(CGF &amp; under-actuation)</b>	Independent	In series	In series	In series	Combined
<b>Installation of CGF</b>	Driving chain	Root joint shaft	Proximal joint shaft	Proximal joint shaft	Driving chain

significant features of TH series hands including TH-1 Hand, TH-2 Hand, TH-GCUA Hand, TH-COSA Hand and TH-3L Hand.

The ‘Fingers’ column gives indicates the number of the fingers in the hand. The next column corresponds to the number of fingers with CGF unit. The under-actuation mechanism, the CGF unit and the relation between the former two are indicated in next three columns, respectively. The final column gives the position of CGF unit installed in the finger. The design of under-actuation mechanism is so crucial that it determines the performance of the hand. The interference between under-actuation mechanism and CGF units must be eliminated. TH-1 Hand does not have this problem, since the under-actuation and changeable forces are not achieved in the same finger. TH-2 Hand balances the two sides of the problem by arranging torsional springs around the root joint shaft. The motor torque is transmitted to the under-actuation mechanism in the finger through the torsional springs. Thus, the relation between under-actuation and CGF is in series. TH-GCUA and TH-COSA adopt the same universal CGF unit design proposed by TH-2. However, this universal design, very simple though, has certain limitations. The grasping forces of the whole finger are limited by the torsional springs. If larger grasping forces are required, it needs far more room to fix larger torsional springs. The whole inter space of the finger is not utilized completely because the springs are separated with under-actuation mechanism. Thus, CGF units should be integrated into the finger in the process of under-actuation design to achieve larger changeable grasping forces. CGF unit is designed form the beginning of the TH-3L Hand design. It is hard to distinguish the CGF unit from the under-actuation mechanism. The two are combined and arranged in the driving chain. The effect of CGF unit is more direct thereby larger grasping forces are possible.

## 4 Analysis of the CGF Unit

### 4.1 Finger Static Model

In this part, a theoretical analysis of grasping forces of on a two-DOF finger to illustrate the effect of CGF unit. Consider a situation both phalanges are in contact with the object. A static model is built to obtain the forces. Figure 6 shows the architecture of the finger where

$O_1$  and  $O_2$  are the centers of the proximal joint and the distal joint, respectively  
 $l_1$  and  $l_2$  are the lengths the proximal phalanx and the distal phalanx, respectively  
 $a, b, d,$  and  $e$  are the lengths of the links,

$c$  is the initial distance between the two sliders,

$f$  is the distance between the center of the slider and  $O_1O_2$ ,

$f_1$  and  $f_2$  are the grasping force on the proximal phalanx and the distal phalanx, respectively,

$h_1$  and  $h_2$  are the force arms of  $f_1$  and  $f_2$ , respectively

$\theta_a$  is the rotating angle of the lowest link with respect to the base,

$\theta_1$  and  $\theta_2$  are the rotating angles of the proximal phalanx and the distal phalanx, respectively,

$T_a$  is the torque of the motor.

According to the principle of virtual work, one obtains

$$\mathbf{t}^T \boldsymbol{\omega}_a = \mathbf{f}^T \mathbf{V} \quad (1)$$

where  $\mathbf{t}$  is the input torque vector exerted by the motor and the under-actuation spring,  $\mathbf{f}$  is the grasping force vector on the two phalanges,  $\boldsymbol{\omega}_a$  is the corresponding rotating velocity vector,  $\mathbf{V}$  is the velocity vector of contact points.

One gets

$$\mathbf{V} = \mathbf{J}_v \dot{\boldsymbol{\theta}} = \begin{pmatrix} h_1 & 0 \\ l_1 \cos \theta_2 + h_2 & h_2 \end{pmatrix} \begin{pmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \end{pmatrix} \quad (2)$$

where  $\mathbf{J}_v$  is expressed as:

$$\mathbf{J}_v = \begin{pmatrix} h_1 & 0 \\ l_1 \cos \theta_2 + h_2 & h_2 \end{pmatrix} \quad (3)$$

Let

$$\boldsymbol{\omega}_a = \mathbf{J}_\omega \dot{\boldsymbol{\theta}} \quad (4)$$

The Jacobian matrix  $\mathbf{J}_\omega$  determines the relation between  $\boldsymbol{\omega}_a$  and  $\dot{\boldsymbol{\theta}}$ . To establish  $\mathbf{J}_\omega$ , velocities of the two sliders are expressed as:

$$v = a(\dot{\theta}_a - \dot{\theta}_1) \left[ \sin\left(\frac{\pi}{2} + \theta_1 - \theta_a\right) - \frac{f}{b} \cos\left(\frac{\pi}{2} + \theta_1 - \theta_a\right) + \frac{a}{2b} \sin 2\left(\frac{\pi}{2} + \theta_1 - \theta_a\right) \right] \quad (5)$$

$$v = e \dot{\theta}_2 \left[ \sin\left(\frac{\pi}{2} + \theta_2\right) - \frac{f}{d} \cos\left(\frac{\pi}{2} + \theta_2\right) + \frac{e}{2d} \sin 2\left(\frac{\pi}{2} + \theta_2\right) \right] \quad (6)$$

The velocities of the two sliders are the same. Solve eq. (5) and eq. (6), one can obtain  $\mathbf{J}_\omega$ . Since  $|J_v| = h_1 h_2 \neq 0$ , finally one gets

$$\mathbf{f}^T = \mathbf{t}^T \mathbf{J}_\omega \mathbf{J}_v^{-1} \quad (7)$$

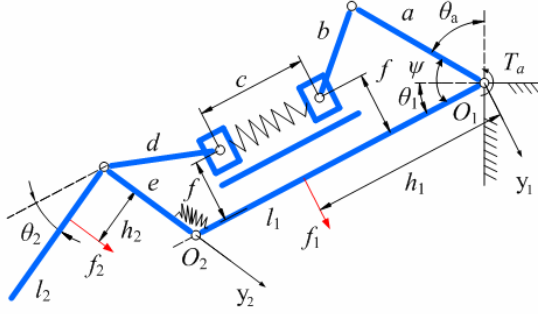


Fig. 4. Architecture of the finger

which is written as

$$\mathbf{f} = \begin{pmatrix} \frac{T_a}{h_1} - \frac{ABT_a + AT_s}{h_1 h_2} \\ \frac{BT_a + T_s}{h_2} \end{pmatrix} \quad (8)$$

In eq. (8),

$$A = l_1 \cos \theta_2 + h_2 \quad (9)$$

$$B = \frac{e \cos \theta_2 + \frac{f}{d} \sin \theta_2 - \frac{e}{2d} \sin 2\theta_2}{a \sin \psi - \frac{f}{b} \cos \psi + \frac{a}{2b} \sin 2\psi} \quad (10)$$

The angle between link  $a$  and the proximal phalanx is expressed as:

$$\psi = \frac{\pi}{2} + \theta_1 - \theta_a \quad (11)$$

## 4.2 Changeable Grasping Forces Exerted by CGF Unit

Now effect of CGF unit is added into the force analysis. The forces produced by the CGF spring is equated to torque imposed on joints to simplify the analysis:

$$T_e = f \cdot f_e = kf \Delta c \quad (12)$$

where  $k$  is the stiffness coefficient of the CGF spring. The torque  $T_e$  pushes the distal phalanx forward while it resists the rotation of the link  $a$ . The motor torque increases to balance the  $T_e$ . The incremental magnitude of the motor torque equals to  $T_e$ . In the finger design, the CGF spring will be much stronger than the under-actuation spring. Hence the equivalent torque of the under-actuation spring is ignored. Then it is obtained that:

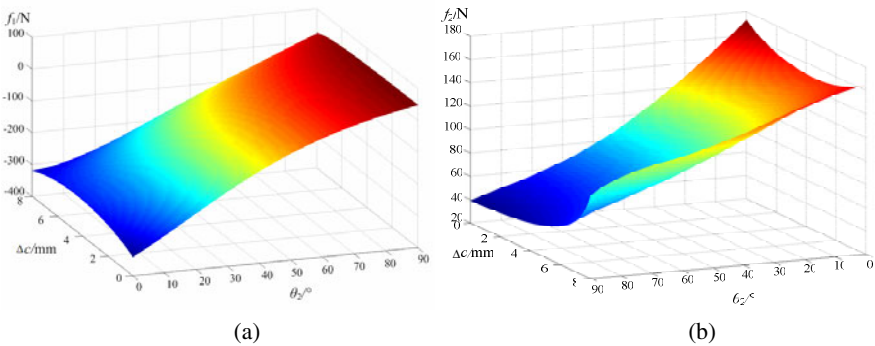
$$\mathbf{f} = \begin{pmatrix} \frac{h_2 - AB}{h_1 h_2} T_a + \frac{h_2 - A(B+1)}{h_1 h_2} T_e \\ \frac{BT_a}{h_2} + \frac{(1+B)}{h_2} T_e \end{pmatrix} \quad (13)$$

The total changeable forces on both phalanges are given by eq. (13). It should be noted that there are two types of changeable grasping forces before further analysis. One type is called total changeable force, since the forces are exerted by both and CGF unit. The other, existing when the motor has stopped, is only generated by the CGF unit called elastic changeable force. The distribution and formation of elastic grasping force are dictated by three factors. The first factor is the configuration of the finger. The second is the mechanical design. The magnitude of the forces is determined by the input torque including the motor torque and the CGF torque. As addressed above, the CGF torque is determined by the motor either. The motor controls the deformation of CGF spring. That is how the changeable forces can be controlled. The maximum of the elastic forces are limited by the CGF spring stiffness and the distance between the two sliders in the CGF unit. The geometric parameters are shown in Table 2 in mm. The motor torque is given by 2 Nm. The stiffness of CGF spring is set to be 1.5 N/mm.

**Table 2.** Geometric parameters

$a$	$b$	$c$	$d$	$e$	$f$	$l_1$	$l_2$
9	28	25	15	6	4	56	35

In Fig. 5 (a) and (b), total changeable forces are expressed as function of  $\Delta c$  and  $\theta_2$ . To reduce the number of variables,  $h_1$  is given by 30 mm and  $h_2$  is given by 15 mm which means contacting points on the phalanges are fixed. Note that  $\theta_1$  is absent in the final expression of total changeable force.



**Fig. 5.** Total changeable forces as function of  $\Delta c$  and  $\theta_2$ . (a)  $f_1$ ; (b)  $f_2$

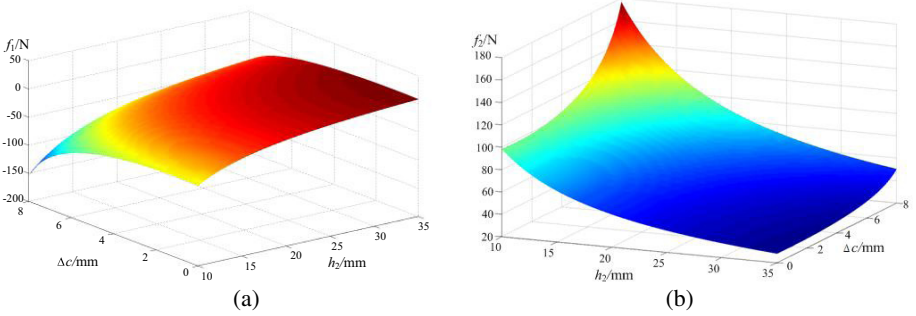
The reason is that it describes the rotating of the finger about the base before any contact is made with the object. The position status of distal phalanx is in primary place. The configuration influences the forces by  $\theta_2$ . The deformation of CGF spring is



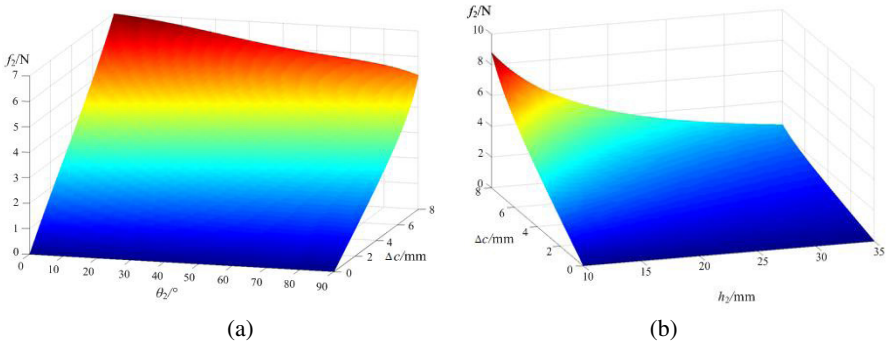
determined by  $\Delta c$ , controlled by the motor. It is illustrated that total changeable forces are mainly influenced by  $\theta_2$  more than  $\Delta c$ . The total changeable forces  $f_1$  rise as  $\theta_2$  increase. On the other hand, effect of  $\Delta c$  on  $f_2$  is more apparent than on  $f_1$ . Increasing of  $\theta_2$  also leads to increasing of  $f_2$ .

In Fig. 6 (a) and (b), total changeable forces are expressed as function of  $\Delta c$  and  $h_2$ . Here  $\theta_1$  is given by 60 degrees therefore a certain grasping configuration is considered. Since  $h_1$  only influence the magnitude of the forces, it is given by 30 mm. The figures show that if contacting points are regarded un-fixed,  $\Delta c$  impacts the total changeable forces apparently. Increasing of  $\Delta c$  will cause declining of  $f_1$ . It is the opposite case with  $f_2$ . The  $f_1$  reaches the minimum point when  $\Delta c$  reaches the upper limit and  $h_2$  reaches its lower limit. In the same condition, the  $f_2$  is at the maximum point.

Now consider only elastic forces remained on the finger when the motor stops. The elastic force  $f_2$  are also expressed as function of  $\Delta c$  and  $\theta_2$  or  $\Delta c$  and  $h_2$  shown in Fig. 7(a) and (b).



**Fig. 6.** Total changeable forces as function of  $\Delta c$  and  $h_2$ . (a)  $f_1$ ; (b)  $f_2$

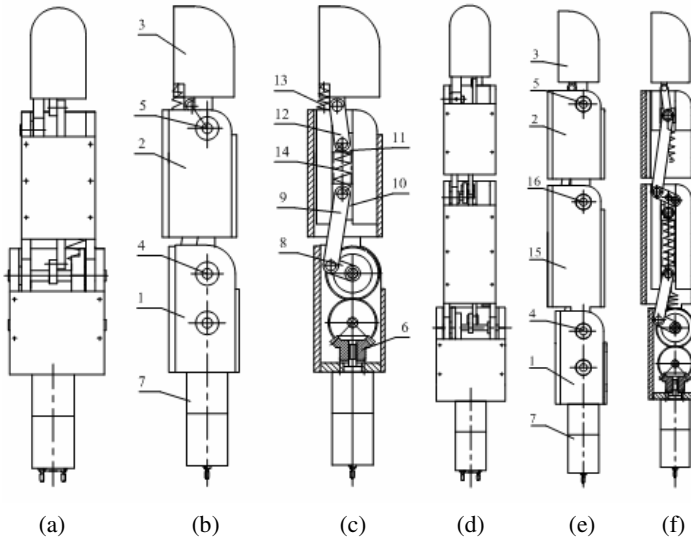


**Fig. 7.** Elastic force  $f_2$ . (a)  $f_2$  as function of  $\Delta c$  and  $\theta_2$ ; (b)  $f_2$  as function of  $\Delta c$  and  $h_2$

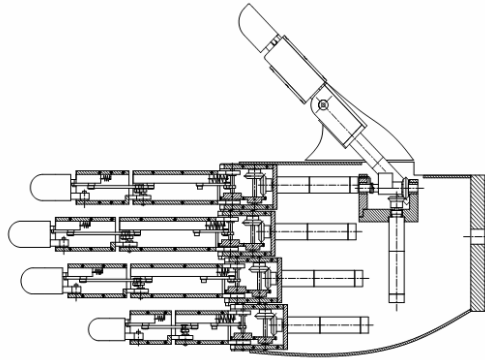
Given by the geometric parameters in Table 2, the elastic force  $f_1$  is negative in all conditions, so that no force will remain on the proximal phalanx virtually. In contrast, the elastic force  $f_2$  is always positive. As a conclusion, CGF unit realize effective remaining changeable force on the distal phalanx without motor torque. A two-phalanx grasp is not always required. However, in most common situation, the terminal phalanx grasp i.e. the distal-phalanx grasp here is sufficient in a whole hand grasp. It is shown that the magnitude of elastic force is dictated by  $\Delta c$ . When  $\theta_2$  increases or  $h_2$  declines, the elastic force  $f_2$  will be larger. The maximum of elastic force is near to 10N, which is sufficient for grasping middle things.

## 5 Design of TH-3L Hand

The designs of the TH-3L fingers are shown in Fig. 8. The TH-3L Hand design is shown in Fig. 9. The motor torque is transmitted to the proximal joint shaft via the gears transmission mechanism. Then the move is transmitted via linkage CGF unit to the distal phalanx. The 3-DOF finger is designed in accordance on the 2-DOF finger adopting the same CGF unit. It is noted that CGF unit is only integrated into the proximal phalanx of the 3-DOF finger. A shared link connects the CGF unit to the upper linkage under-actuation mechanism.



**Fig. 8.** Mechanical design of TH-3L fingers. (a)Front view of 2-DOF finger; (b)Side view 2-DOF finger; (c)Side cutaway view 2-DOF finger; (d)Front view of 3-DOF finger; (e)Side view 3-DOF finger; (f)Side cutaway view 3-DOF finger. 1-base; 2-proximal phalanx; 3-distal phalanx; 4-proximal joint; 5-distal joint; 6-gears transmission mechanism; 7-motor; 8-first link; 9-second link; 10-first slider; 11-second slider; 12-third link; 13-under-actuation spring; 14-CGF spring; 15-middle phalanx; 16-middle joint.

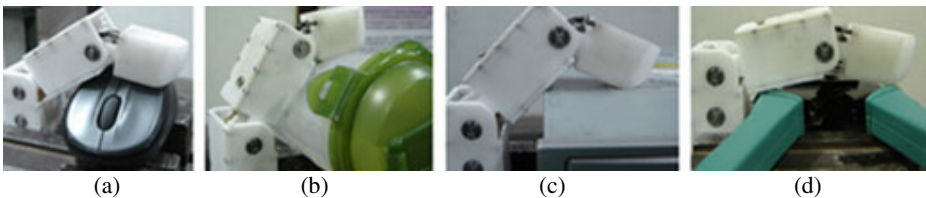


**Fig. 9.** Mechanical design of TH-3L Hand

The 2-DOF finger and the 3-DOF finger are mounted in a five-fingered hand, named TH-3. The whole hand has 15 DOFs using 6 motors. Changeable grasping forces are achieved on all five fingers with the modular CGF units. Except for the thumb, the other four fingers have the same structure equipped with one motor actuating three joints. A motor is embedded in the palm to actuate the thumb to swing around the side of the hand.

## 6 Experiments of TH-3L Finger

To verify the CGF unit, a prototype of 2-DOF finger was manufactured and relating experiments were conducted. The first goal is to test the self-adaptation to objects. A series of grasp experiments are shown in Fig. 10 in which various targets were selected including cuboids, cylinders and irregular-shape objects. The finger performed a firm grasp, two phalanges enveloping the object. The design is successful in the terms of under-actuation, no interference form CGF unit generated.



**Fig. 10.** Grasps of different objects by TH-3L finger

After grasping, total changeable forces are exerted by CGF unit. Experiments were designed to testify the CGF effect. A Motorman six-joint robot was used providing a fixed prop hung with a weight beam used to detect the change of the grasping force. A light object here is hung on the weight beam. Firstly, the finger contacts the object and grasps it without large load. Then the motor torque rises so that changeable forces are

generated while the configuration of the finger remains. The display on the weight beam will show the change of the grasping forces as a sum. The initial display of the weight beam was about 0.1 kg (10 N approximately) when only the distal phalanx contacts the object. Then it reached 0.2 kg (20 N approximately) after a slight increment of the motor torque. Total changeable force are proved to exist.

## 7 Conclusion

Research work of CGF unit design in TH series hands was concluded. Based on the conclusion, a novel design was proposed to combine CGF unit with under-actuation mechanism, changing traditional approaches. Comparing to proceeding hands, TH-3L Hand with the novel linkage CGF unit has many advantages in the context of changeable forces. A large changeable force is obtained with compact structure. The CGF unit is analyzed in detail to know about the total changeable forces and elastic forces. Relation between parameters and the changeable forces are given. The changeable force is kept effectively on the distal phalanx verified by theoretical analysis, solving the problem addressed in the beginning if the paper. As a result, TH-3L Hand is provided with changeable grasping forces. The linkage CGF unit for TH-3L Hand exposes a design approach combining CGF unit and under-actuation. More CGF units can be achieved in this approach to be used in robotic hands.

**Acknowledgements.** This research was supported by Hi-Tech R&D Program of China (No. 2007AA04Z258), National Natural Science Foundation of China (No. 50905093), Tsinghua Basic Research Foundation (No.JC2007009) and the foundation of Key Laboratory for Advanced Material Processing Technology, Ministry of Education, P.R. China (No.2008011).

## References

1. Laliberte, T., Gosselin, C.: Simulation and Design of Under-actuated Mechanical Hands. *J. Mech. Mach. Theory* 33(1/2), 39–57 (1998)
2. Laliberte, T., Birglen, L., Gosselin, C.: Under-actuation in robotic grasping hands. *J. Machine Intelligence & Robotic Control* 4(3), 1–11 (2002)
3. Akin, D.L., Carignan, C.R., Foster, A.W.: Development of a Four-Fingered Dexterous Robot End Effector For Space Operations. In: *IEEE Inter. Conf. on Rob. and Aut.*, Washington, DC, pp. 2302–2308 (2002)
4. Biagiotti, L., Melchiorri, C., Vassura, G.: Control of a Robotic Gripper for Grasping Objects in No-Gravity Conditions. In: *IEEE Inter. Conf. on Rob. and Aut.*, Seoul, Korea, pp. 1427–1432 (2001)
5. Figliolini, G., Ceccarelli, M.: A Novel Articulated Mechanism Mimicking the Motion of Index Fingers. *J. Robotica* 20, 13–22 (2002)
6. Licheng, W., Ceccarelli, M.: A Numerical Simulation for Design and Operation of an Under-actuated Finger Mechanism for LARM Hand. *J. Mechanics Based Design of Structure and Machines* 37(1), 86–112 (2009)
7. Zhang, W., Che, D., Liu, H., et al.: Super Under-actuated Multi-fingered Mechanical Hand with Modular Self-adaptive Gear-rack Mechanism. *J. Industrial Robot: An International Journal* 36(3), 255–262 (2009)

8. James, F.M.: Mechanical Hand. US Patent: No. 3694021 (1972)
9. Sukhan, L.: Artificial Dexterous Hand. US Patent: No. 4955918 (1990)
10. Nathan, T.: Methods and Apparatus for Mechanically Intelligent Grasping. US Patent: No. 4957320 (1990)
11. Dechev, N., Cleghorn, W.L., Naumann, S.: Multiple finger, passive adaptive grasp prost. *J. Mech. Mach. Theory* 36, 1157–1173 (2001)
12. Zhang, W., Chen, Q., Sun, Z., et al.: Under-actuated passive adaptive grasp humanoid robot hand with control of grasping force. In: *IEEE Inter. Conf. on Rob. and Aut.*, Taipei, Taiwan, pp. 696–701 (2003)

# Efficient Grasp Planning with Reachability Analysis

Zhixing Xue and Ruediger Dillmann\*

Forschungszentrum Informatik,  
Haid-und-Neu Str. 10-14, 76131 Karlsruhe, Germany  
{xue,dillmann}@fzi.de  
<http://www.fzi.de/ids>

**Abstract.** Grasping can be seen as two steps: placing the hand at a grasping pose and closing the fingers. In this paper, we introduce an efficient algorithm for grasping pose generation. Depend on the hand kinematic, boxes of different sizes are sampled. The reachability for grasping is represented by the information, from where the hand can grasp the box firmly. These boxes represent real objects, which at run-time will be decomposed into such boxes, so that the grasping poses for the real object can be generated. Concrete grasps at a grasping pose will be further checked for its grasp quality. Real experiments with two different robotic hands show the efficiency and feasibility of our method.

## 1 Introduction

In the last decades, many robotic hands have been developed. To bring these hands successfully from the research laboratory to household or industrial environments, grasp planning system are needed, which are able to compute feasible and stable grasps for robotic hands automatically within less than a few seconds. Grasp planning has become an area of great interest in robotics with two main properties: high-dimensional configuration space and stability analysis. To perform a planned grasp, not only the finger joint positions, but also the Cartesian pose of the robotic hand and its corresponding joint configuration of the robotic arm have to be considered. A brute force search in this high-dimensional configuration space would be intractable. In motion planning problem, the evaluation of a configuration in this high-dimensional space is determined by the results of collision checking, whereas it is determined by the stability analysis by the grasp planning. Not the configuration itself but the contact points between the object and the robot at this configuration will be used to check if it is a stable grasp, using methods like the force closure property [1].

The crucial part of a grasp planning algorithm is to find out heuristics to reduce the high-dimensional configuration space. The hand's internal degrees of

---

\* The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 216239.

freedom and the Cartesian pose of the hand can be treated separately. Wren and Fisher [2] proposed to use hand preshapes to reduce the finger configuration space. Grasping is performed by placing the hand into a Cartesian pose to encompass the object and then closing the fingers until they have contact with the object. This high-dimensional finger configuration space can also be described in a low dimensional space with the Eigengrasps [3]. The Cartesian pose of the robotic hand should be chosen where the object can be grasped firmly. Many approaches work in an object-centered way where the object is decomposed into smaller pieces, from which the hand pose is determined. In grasp simulator GraspIt! [4], the object is decomposed into shape primitives such as boxes, spheres, cylinders and cones, from which starting positions and approaching direction towards the object for the robotic hand are generated. In the simulation, the hand within a predefined preshape starts to move from one of the starting position along an approaching direction and try to grasp the object by closing the fingers. Other object decomposition and representation methods have been proposed, such as Superquadrics [5] and minimum volume bounding boxes [6].

The grasping methods mentioned above can be improved, if the *reachability* of finger and hand are taken into account, as will be shown in this paper. During the generation of grasping directions, a problem which is still ignored is that the starting positions and the grasping directions are object dependent, without consideration of the kinematic information of different robotic hands. Instead of a possible *grasping pose*, from which the object can be grasped firmly, a starting position and an approaching direction are given. The hand needs to try to move and close many times until it can reach the object in the simulation. Due to the geometrical approximation, the generated starting positions do not cover all possible grasping poses. The computed results are saved in grasp databases, which is accessed during the real execution to find a feasible grasp which is reachable by the robotic arm [7, 8].

In this paper, a novel grasp planning algorithm with reachability analysis is proposed. Given the kinematic information of the robotic hand, its preshapes are predefined. For each of the preshape, the swept volumes of the finger links during flexing are generated to represent the *hand reachability* with this preshape. Based on this hand reachability analysis, the possible graspable boxes of different sizes are generated. The boxes are placed into discretized Cartesian poses within the graspable region of the hand, where the collision between the faces of the box and the swept volume of the fingers are checked to find out grasping poses that are suitable for this box. This way, the *grasping reachability* for each of graspable box is computed. It depends only on hand kinematic and can be computed off-line. In the on-line phase the object that is to be grasped is decomposed hierarchically into graspable boxes, from which the grasping poses for the robotic hand can be computed directly. With the reachability analysis and the graspable boxes, the object and its grasping poses are connected together.

## 2 Reachability Analysis

In this section, the reachability of finger, hand and grasping are analyzed. The grasping reachability is derived from the finger reachability and hand reachability. It only depends on the robot’s hand kinematic information and is independent of objects that can be grasped by the hand. Without considering a concrete object, we want to concentrate on the volumes that can be reached and touched by a single finger and the whole hand. This leads naturally to the computation of the swept volumes.

### 2.1 Finger Reachability

The finger model we consider is anthropomorphic, similar like the human finger. The rotation of the beginning joint causes abduct and adduct movements. The following joints flexes the finger in a same plane. As many robotic fingers are built this way, the following analysis can be applied to these robotic hands. As reviewed in the introduction section, the grasping process can be decomposed into two sub-processes: placing hand preshape and finger closing. In this paper, the hand preshapes are defined so that the fingers have different initial joints. During the finger closing process, the beginning joints will not be moved and only the following joints will be rotated to grasp the object. It should be noticed that with different joint velocities during the interpolation, different contact points on the object surface can be touched as reported in our previous work [9].

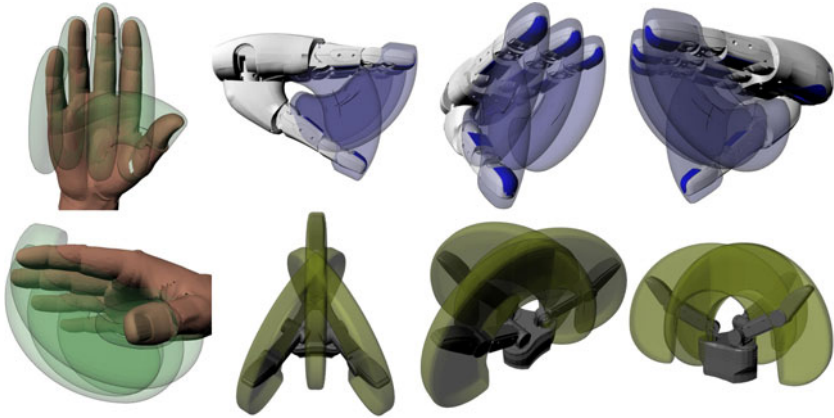
Because the last finger link at the end of the kinematic chain can reach the largest region during the finger closing, we use the swept volume of the last finger link with respect to all of the finger joints, except the beginning joint fixed by the preshape, to represent the reachability of the finger. A point in this swept volume or an object intersecting with this swept volume indicates that the point or the object can be reached by the finger. The recursive generation of the finger swept volumes of a human hand model from GraspIt! [10], the Schunk Anthropomorphic Hand (SAH) with four fingers and Schunk Dexterous Hand (SDH) with three fingers are shown in Fig. 1.

### 2.2 Hand Reachability

The reachability of the robotic hand is represented by the reachabilities of all its fingers. If an object intersects with some finger last links’ swept volumes, it can be reached by the hand using these fingers. It means that it is possible to grasp the object with these fingers. Otherwise, if an object does not touch any swept volumes, it is not reachable for the hand at this Cartesian pose and no grasp planning needs to be performed. This is an efficient method to check if an object is reachable or not with only a few collision checking between the swept volumes and the object.

Our algorithm is based on this reachability checking. Only if the object collides with swept volumes of more than two fingers, it will be tried to grasp the object. There are two special cases where grasps of small objects are lost. One is that





**Fig. 1.** Recursively generated finger links' swept volumes of different hand models illustrated with transparent colors. Left: swept volumes of a human hand model. Upper right: swept volumes of SAH. Lower right: swept volumes of SDH. These swept volumes are used to represent the reachability of the hand.

the object can be grasped by only one finger. The other is that a small object is near the palm, and does not collide with the swept volumes of the last finger links, but can be grasped with other finger links. These two special cases are ignored by our algorithm.

### 2.3 Graspable Boxes

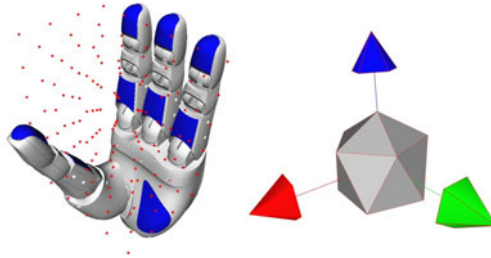
Although whether an object is reachable for the hand can be easily checked using our method, the check needs to be performed for each object, after its Cartesian pose is available. Despite the reachability of the object for the hand, it is still not sure if the object can be firmly grasped. We want to find out directly all possible grasping poses for all possible objects. To do this with reasonable complexity, we discretized the continuous six-dimensional Cartesian space into discrete sample poses and approximate the possible grasping objects with its sampled bounding boxes, which we call *graspable boxes*. The main idea is to test all sampled grasping poses relative to each of the graspable boxes of a given hand, find out feasible grasping poses and save the results for later use. At run-time, the object will be approximated by its bounding box and decomposed into smaller bounding boxes. Grasping poses of the graspable boxes similar to these bounding boxes will be considered as grasping poses for the object. Because the graspable boxes can be grasped by the hand, it is highly likely that the object or object parts can also be grasped from these generated grasping poses. The provided grasping poses will be further checked for the reachability with the arm, the stability of the grasp and so on.

**Approximated Graspable Boxes.** To grasp objects of all possible sizes, we need to find a way to approximate the different sizes. Box is one of the simplest shape primitives in three dimensional space. The bounding box of an object describes the volume that completely contains the object geometry, which is often used for collision checking. A box has six faces and eight vertexes. Width, height and depth of a box determine the size of a box. Because of its simplicity, we analyze how each box can be grasped with the robotic hand instead of concrete object geometries. It is based on the assumption that if the bounding box of an object can be grasped, it is likely that the object itself can also be grasped. In fact, because the bounding box encloses the object, a preshape that does not collide with the bounding box can not touch the object inside.

The size of an object, which can be grasped with a given robotic hand is determined by the kinematic of the hand. The maximum opening distance of the hand between two spread fingertips is the maximum length a graspable box can have. The diameter of the fingertip is also an important metric for grasping. It is still very difficult to grasp and stabilize objects that are smaller than the robot fingertip with the robotic hands. It is used as a parameter to sample graspable boxes with different sizes and to sample the Cartesian poses of the graspable boxes in the graspable region of the hand. Given the opening distance of the hand  $D$  and the diameter of the fingertip  $r$ , the sampled graspable boxes are generated. Every side is sampled with the ratio between the opening distance and the diameter of fingertip  $n = \lfloor \frac{D}{r} \rfloor$  steps, resulting in a total  $n^3$  boxes. The length of width, height and depth is sampled  $l = i \times \frac{D}{n}, i \in [1..n]$ . A human hand with opening distance of 200 mm (between fingertips of thumb and ring finger) and fingertip diameter of 15 mm as example,  $n = \lfloor 200/15 \rfloor \approx 14$  different sizes will be used to generate the graspable boxes. Other discretization step instead of the fingertip diameter can also be used to sample the graspable boxes.

**Discretized Cartesian Poses.** After the graspable boxes are determined, grasping poses from which the box can be grasped have to be found. A box has to be necessarily reachable for the hand to be grasped. Because the possible grasping poses are continuous in the Cartesian space, it also needs to be discretized. With the analysis of the hand reachability introduced in the last subsection, we compute the bounding box of the swept volumes of the last finger links. If the center of a graspable box lies in this bounding box, it very likely intersects with the swept volumes and is reachable with the fingers. The positions of the possible grasping poses are sampled in this bounding box. At each position, uniformly distributed orientations are generated, which together with the positions build the possible grasping poses in Cartesian space for the graspable boxes. Here, the diameter of the fingertip is used again as step size to sample the positions in the bounding boxes of the finger links' swept volumes. Later during grasping, if a grasping pose is not valid due to e.g. collision with the environment, the next grasping poses near it will be tested with the distance of a fingertip diameter to it. This allows efficient grasping in complex environment.

After the positions are generated in the bounding box of the reachable region for the hand, shown as the red dots in Fig. 2, the same uniformly distributed orientations at each position are computed to generate the Cartesian grasping poses. 12 equally distributed points on a sphere are chosen to determine the first rotational axis of the orientation. It is generated by an axis from the origin to one of the vertexes of a regular icosahedron, which is located at the origin. Around this axis, it is then rotated in a fixed step size. Each time, a sampled orientation is created. With  $30^\circ$  as the rotation step size, there are totally 144 orientations at each generated position. One of the orientations generated this way is presented in Fig. 2.



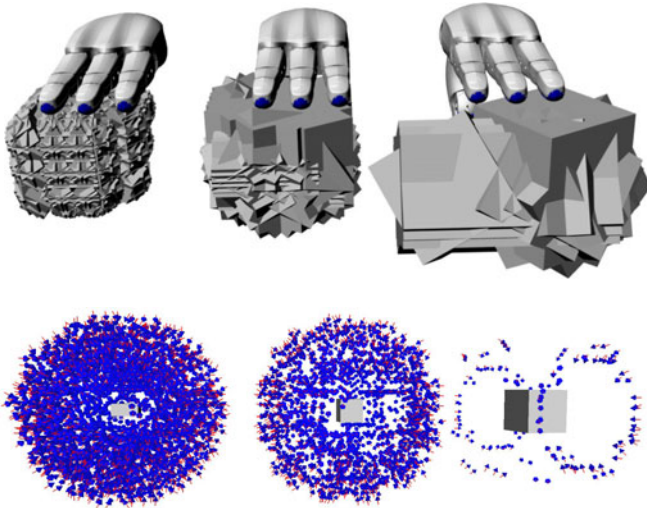
**Fig. 2.** The reachable region of the robotic hand is equally sampled. Left: red dots represent the sampled positions with the diameter of the fingertip as step size. Right: 144 Orientations at each position are generated using the 12 vertexes of a regular icosahedron as the first rotational axis.

## 2.4 Grasping Reachability

With the reachability of the hand represented by the finger links' swept volumes for one hand preshape and the generated reachable graspable boxes, grasping reachability of the hand is the computational results, a graspable box from which grasping poses can be grasped. In order to grasp a graspable box, and to grasp the object represented by it, some conditions must be fulfilled:

- The hand preshape may not collide with the graspable box, so that the grasp later can be performed.
- At least two fingers must be used for grasping.
- More than one face of the graspable box have to be reachable by the fingers. Otherwise, if only one face can be reached, the object can only be pushed or pulled, but not grasped.
- Some faces of the graspable box are not reachable during finger closing. These hidden faces should not be considered by the reachability computation.
- The swept volumes of the finger links represent the possible reachable volume of the finger link. The graspable box represents the real object to be grasped. The real grasp with concrete finger positions and object geometry still need to be checked.

The points listed above are performed to compute the grasping reachability. Given a graspable box with width  $w$ , height  $h$  and depth  $d$  at origin, the hand is placed at one of the generated Cartesian grasping pose  $(p, \phi)$  relative to the box, the collision checking are performed twice. If the graspable box does not collide with the hand within predefined preshape, the collision between the swept volumes of the finger  $F_i, i \in [1..n_f]$  and the six faces of the box  $f_j, j \in [1..6]$  is checked, where  $n_f$  denotes the fingers number of the hand. If they collide, the face is reachable by the finger and it is possible that the object in this graspable box can be reached by this finger from the direction of the face. All checking results are saved in a look-up table with each element expressed as a tuple  $[p, \phi, w, h, d, \mathbf{F}_{\text{face}}, \mathbf{F}_{\text{finger}}]$ , which we call the *reachability data*. Fingers that collide with the graspable box are described in a set  $\mathbf{F}_{\text{finger}} = \{F_i\}$  and the touched faces in another set  $\mathbf{F}_{\text{face}} = \{f_j\}$ . This collision checking depends only on hand kinematic. The graspable boxes, all discretized grasping poses and the swept volumes of the last finger links are generated from the hand kinematic. So given a robotic hand that should be used, the whole reachability checking needs to be computed only once off-line. To cut off the faces that are unreachable for finger closing, the angle between the eight vertexes of the box and the first flexing joint of the finger is computed. The faces intersecting at the vertex with the minimum positive angle could be reached first during the finger closing process and will be considered as reachable faces for the collision checking. If the axis of the box is parallel to the first flexing joint axis of the finger, there are two



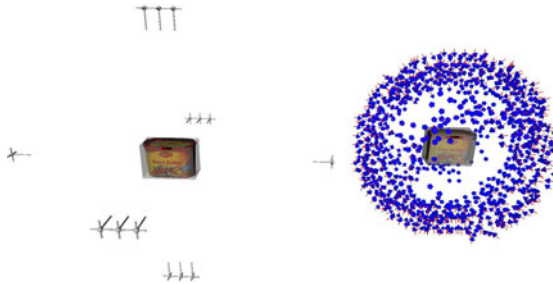
**Fig. 3.** Different sizes of graspable boxes are generated which will be placed at discretized Cartesian poses of the graspable region of the hand. If more than two faces of the box can be reached by at least two fingers, the box could possibly be grasped by the hand and so the object lies within it. First row shows three graspable boxes with different sizes at feasible poses. Second row shows the resulted grasping poses in the coordinate of the graspable box.

or four vertexes with the same minimum positive angle. 3, 4 or 5 faces can be found as reachable faces. The other faces are hidden faces with which the swept volume could collide with, but which are not reachable during the finger closing. Three graspable boxes of the SAH generated this way are shown in Fig. 3.

## 2.5 Real Object Approximation Using Graspable Boxes

With the knowledge about the reachability of the hand, next step is to approximate and decompose the real object using bounding boxes, find out the similar graspable boxes used by the grasping reachability analysis and derive the grasping poses from it, as presented in Fig. 4. The object approximation using bounding boxes is inspired by the work of Huebner et al. [6]. They approximate the object geometry at first with a minimum volume bounding box [11]. To avoid computation in three dimensional space, an optimal cut is found in the three 2D planes perpendicular to the three axes with point clouds projected onto the three sides of the bounding box. Bounding boxes of the decomposed point clouds are separately computed and the algorithm decomposes the object in a hierarchical way, until some stop criteria are fulfilled.

We have implemented this algorithm for the hierarchical object decomposition with minimum volume bounding boxes and extended it with checking of invalid faces. The faces of each bounding box are checked if they can be reached by the finger due to the collisions with other decomposed boxes in the same decomposition level. For example, if two decomposed bounding boxes lie besides each other, the two neighbored faces of the two bounding boxes will be eliminated after this checking, so that grasping from inside an object is avoided. This way, the set of invalid faces  $F_{\text{invalid}}$  that should not be touched by the hand can be created. For each of the resulting bounding boxes of the real object, a previously generated graspable box larger than it in all three directions with minimum volume among the graspable boxes is found. Given a size length  $l'$  of the bounding



**Fig. 4.** Left: generated starting positions and approach directions by grasp planner in GraspIt!. Results are only object dependent, the hand needs to move from a starting position along the approach direction to find the final grasping pose. Right: generated grasping poses by our method for the same object with the SAH as shown in Fig. 3. The results are directly final grasping poses, which depend on both the object and the used hand.

box, the next length used to generate the graspable box is  $[l' \times \frac{n}{D}] \times \frac{D}{n}$ . After a corresponding graspable box with size  $[w', h', d']$  for a bounding box is found, the grasping poses for the object can be accessed directly from the reachability data:  $[p, \phi, w', h', d', \mathbf{F}_{\text{face}}, \mathbf{F}_{\text{finger}}]$ . Only the grasping poses  $(p, \phi)$  at which the fingers do not collide with the invalid faces will be used for grasping, i.e.  $\mathbf{F}_{\text{invalid}} \cap \mathbf{F}_{\text{face}} = \emptyset$ . If there is no graspable box which can completely contain a bounding box of the object, the object is too big for the robotic hand to be grasped. Graspable boxes for smaller bounding boxes decomposed from the big bounding box can be found, with which it will be tried to grasp the smaller parts of the object. The grasping poses are in bounding boxes' coordinates. After transforming into the object's coordinate, the generated grasping poses need to be checked to find out grasping poses that are also reachable by the robotic arm involving the arm's inverse kinematic algorithm.

### 3 Grasp Planning with Reachability Analysis

With the reachability analysis introduced in the last section, a main question for grasping planning is answered: where should the hand be placed to grasp an object. Our method uses heavily bounding volumes approximation to find the grasping poses, where the hand can reach the object, can grasp its bounding box with more than one finger, where the hand can touch more than one side of its bounding box, and the hand can also be moved there using the robotic arm.

If information such as 3D geometry, weight and surface friction property of an object are available, physical stability of a grasp can be exactly computed. Many work has been done in the past to measure the grasp quality for known objects. The 3D geometry model of an object can be decomposed hierarchically to bounding boxes, from which the grasping poses can be computed. With the robotic hand placed at one of the computed grasping poses, fingers close towards the object until collision occurs. The closing motion should be interpolated and the interpolation should stop at the first contact point. This is a problem of continuous collision detection. It can find the first contact point between two moving objects. The continuous collision detection library [12], [13] we use is based on conservative advancement. The distance between two moving objects is computed and the upper motion bound between them estimated. The algorithm computes a time interval, where the motion can be advanced without collision. It works in an iterative way until the distance is below an user defined threshold. Only a few iteration steps are needed. In our previous work [13], swept volumes of finger links were used to improve the efficiency of the continuous collision checking by eliminating colliding pairs that do not collide with each other.

After the contact points between finger links and the object are found, contact forces with Column friction model at the contact points along the contact normals are computed. With the object's center of mass as chosen reference point, torques can also be computed, which together with the grasping forces build the grasping wrench space in six dimensional space. We use the "largest sphere" quality measurement as implemented in GraspIt! to compute the grasp quality.

From each of the grasping poses, the collision checking and the grasp quality computation are performed. The search ends if a grasp with more than 0.01 grasping quality is found. We have intensively tested the planned grasps with real robotic hands and found out that grasps with more than this grasping quality are sufficient stable. The friction coefficient was set to be 0.3. If the fingers are moved to the joint positions with the contact points, the fingers only touch the object and can not apply forces onto it. The forces are optimized as a linear matrix inequalities problem [14] and can be applied using finger joint torque impedance control [7]. For the SDH with tactile sensors on the finger surface, we have defined a penetration depth at the contact points for the soft rubber parts. With the transformation using the contact Jacobi matrix, the fingers are moved towards the object further as at the contact points, so that the contact areas between the finger pads and the object with the desired penetration depth can stabilize the object in the hand.

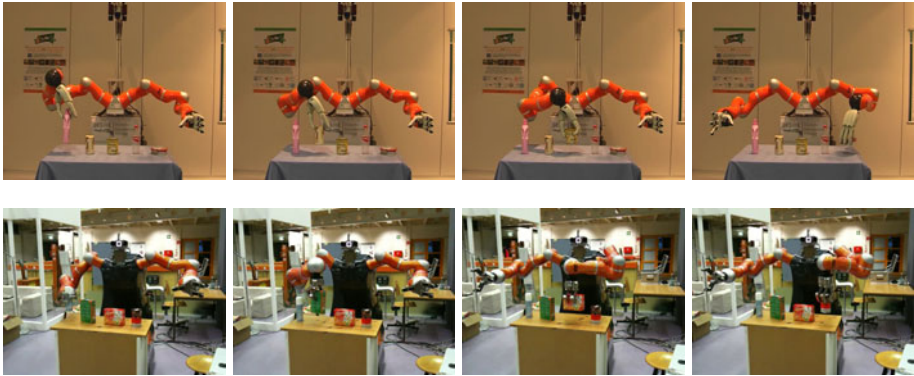
After a high quality grasp is found and the finger joint positions determined, the inverse kinematic algorithm of the robotic arm is used to find the arm joint positions to bring the hand to the grasping pose. Direct after the arm joint positions are computed, self collision checking and collision checking between the robot and the environment are performed. A motion planning algorithm is applied for collision free grasping. Also, the arm impedance control, like the one provided by the KUKA light weighted robotic arm can be used to allow safety physical interaction between the robot and its environment.

## 4 Experimental Results

Three hand models have been tested with the proposed method. The preshapes for the human hand model from GraspIt!, the SAH and the SDH are defined and their swept volumes are recursively generated, as illustrated in Fig. 1. Due to paper limitation, only the experimental results with the SAH are reported here. The diameter of the fingertip is 34 mm and the maximum opening distance is 130 mm. 64 graspable boxes with different sizes are generated. The collision checking between the swept volumes and each faces of the graspable boxes are performed using library SWIFT++ [15]. The time needed was nearly 9 minutes. It needs only to be computed once off-line. The resulted reachability data are saved for later use. In Fig. 3, feasible graspable boxes are shown, which do not collide with the hand in preshape and can be reached by the fingers. The graspable box in the first column has the size of [32.5, 65, 32.5]. It has the most grasping poses (2928) found. The second graspable box with size [65.0, 65.0, 65.0] has 1648 grasping poses. The last graspable box is the largest one with size of [130.0, 130.0, 130.0] and has 128, the least grasping poses.

Fig. 4 compares the grasping pose generation between the automatic planner in GraspIt! and our method. It is shown that the resulted starting position and approach direction by GraspIt! do not provide the final grasping pose, from which the hand can grasp the object directly. Cartesian interpolation of the hand from the starting position along the approach direction is needed to get





**Fig. 5.** Grasping experiments with the SAH and the SDH. The robot computes feasible grasps for both of the left and right hands and executes the pick and place operation. To apply forces onto the object, finger joint torque impedance control of the SAH is used for optimized grasping forces. Tactile sensors of the SDH are used to gain soft contact areas with the object and to check the contact status.

the final grasping pose. Also, it is clearly illustrated that not all directions are covered as in our case.

As explained in Section 3, stable grasps can be found from the generated grasping poses. Actually more than 100 household objects are automatically modeled in our object modeling center [7], [16]. All of the 100 objects can be grasped using both SAH and SDH. Some grasping results are shown in Fig. 5. The robot tries to find grasps with both hands, if it can find a feasible grasp with one of the hand, it will be performed. The movement of the robotic arm is planned using motion planning algorithm. During pick and place operation, the robotic arm is controlled in impedance mode to allow safety interaction.

## 5 Conclusion and Outlook

In this paper, a novel algorithm for grasping poses computation was proposed. The algorithm uses the bounding volumes to generalize the volumes that can be occupied, touched and grasped. Reachability of the finger is represented by the swept volume of the last finger link. Reachability of the hand is represented by the reachabilities of all the fingers. Dependent on hand reachability, bounding boxes with different sizes that could be grasped by the hand are generated and placed at sampled Cartesian poses in the hand reachable space. If more than one face of the box can be reached by at least two fingers, it is possible to grasp the object within this graspable box with the hand. All of the grasping poses generated this way depend only on hand kinematic and need to be computed only once off-line. With this reachability knowledge, the object to be grasped will be decomposed hierarchically at run-time to a few bounding boxes to compute good grasping poses, from which the hand can reach and grasp the object. Experiments with



real robotic hands show that our method is very efficient and provides direct reachable grasping poses.

## References

1. Ferrari, C., Canny, J.: Planning optimal grasps. In: IEEE International Conference on Robotics and Automation (1992)
2. Wren, D., Fisher, R.: Dextrous hand grasping strategies using preshapes and digit trajectories. In: IEEE International Conference on Systems, Man and Cybernetics (1995)
3. Ciocarlie, M., Goldfeder, C., Allen, P.: Dimensionality reduction for hand-independent dexterous robotic grasping. In: International Conference on Intelligent Robots and Systems (2007)
4. Miller, A., Knoop, S., Christensen, H., Allen, P.: Automatic grasp planning using shape primitives. In: IEEE International Conference on Robotics and Automation (2003)
5. Goldfeder, C., Allen, P., Lackner, C., Pelossof, R.: Grasp planning via decomposition trees. In: IEEE International Conference on Robotics and Automation (2007)
6. Huebner, K., Ruthotto, S., Kragic, D.: Minimum volume bounding box decomposition for shape approximation in robot grasping. In: IEEE International Conference on Robotics and Automation (2008)
7. Xue, Z., Kasper, A., Zoellner, J.M., Dillmann, R.: An automatic grasp planning system for service robots. In: 14th International Conference on Advanced Robotics, ICAR (2009)
8. Goldfeder, C., Ciocarlie, M., Dang, H., Allen, P.K.: The columbia grasp database. In: IEEE International Conference on Robotics and Automation (2009)
9. Xue, Z., Zoellner, J.M., Dillmann, R.: Automatic optimal grasp planning based on found contact points. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM (2008)
10. Miller, A., Allen, P.: Graspit! a versatile simulator for robotic grasping. IEEE Robotics & Automation Magazine (2004)
11. Barequet, G., Har-Peled, S.: Efficiently approximating the minimum-volume bounding box of a point set in three dimensions. J. Algorithms (2001)
12. Zhang, X., Redon, S., Lee, M., Kim, Y.J.: Continuous collision detection for articulated models using taylor models and temporal culling. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2007) (2007)
13. Xue, Z., Woerner, P., Zoellner, J.M., Dillmann, R.: Efficient grasp planning using continuous collision detection. In: IEEE International Conference on Mechatronics and Automation, ICMA (2009)
14. Han, L., Trinkle, J., Li, Z.: Grasp analysis as linear matrix inequality problems. IEEE Transactions on Robotics and Automation (2000)
15. Ehmann, S.A., Lin, M.C.: Accurate and fast proximity queries between polyhedra using convex surface decomposition. In: Computer Graphics Forum (2001)
16. KIT ObjectModels Web Database, <http://wwwiaim.ira.uka.de/ObjectModels>

# Development of an Anthropomorphic Prosthetic Hand for Man-Machine Interaction

Nan Li<sup>1</sup>, Li Jiang<sup>1</sup>, Dapeng Yang<sup>1</sup>, Xinqing Wang<sup>1</sup>, Shaowei Fan<sup>1</sup>, and Hong Liu<sup>2</sup>

<sup>1</sup> State Key Laboratory of Robotics and System, Harbin Institute of Technology,  
150001 Harbin, China

<sup>2</sup> Institute of Robotics and Mechatronics, German Aerospace Center,  
82230 Munich, Germany  
nanli0111102@gmail.com

**Abstract.** This paper presents a novel highly integrated prosthetic hand equipped with a man-machine interface for performing good grasping features. The hand has five fingers and each of them can be driven individually by the embedded actuation and control system fitted in the palm. By using the man-machine interface, users can control the prosthetic hand by means of their EMG signals and get grasp force feeling through the sensory feedback system designed based on surface electrical stimulus. At the end of this paper, a grasping trial is presented to verify the ability of interaction between the human body and the prosthetic hand.

**Keywords:** Prosthetic hand, Man-machine interface, Sensory feedback, Embedded control system.

## 1 Introduction

The prosthetic hand technology has made great progress under the promotion of robotics and biotechnology. With respect to the semblance [1], control method [2] and dexterity [3], the prosthetic hand is getting close to the real hand. But it should be noticed that the function of normal hand is mainly divided into two aspects: 1. completing the body action under the control of human brain. 2. sensory feedback from body to human brain. It is an obvious process of bidirectional transmission of information realized via passing bioelectrical signal on neural pathway connecting the brain and the hand together. Unfortunately, such mutual awareness can not be fully implemented between the amputee and prosthetic hand, because of the inexistence of the bidirectional messaging channel.

For solving this problem, many researchers have employed some alternative approaches, such as separately establishing the control channel from human-body to prosthetic hand and the sensory feedback channel from prosthetic hand to human body. Recently, the electromyography (EMG) control is mostly applied on the prosthetic hand, resulted from the mature technology of controlling prosthetic hand by identifying EMG patterns [4]. However, the investigation of sensory feedback has not yet formed a mature approach. Reference [5] tried to obtain the sensory feedback in force through using vibration motor under the CyberHand. The tendon-driven prosthetic hand of University of Tokyo [6] initially achieved tactile feedback by

adopting surface electrical stimulus method [7]. Yet, their control system of the prosthetic hand is based on PC and too large to carry out portably.

In this paper, an intelligent prosthetic hand system named HIT IV hand is proposed to perform man-machine interaction. This prosthetic hand has 5 degree of freedom, 15 rotating joints and can handle all control algorithms in a high integrity control system. The interaction between human body and prosthetic hand is functioned by the discrimination of EMG signals and execution of electrical stimulation on body surface.

This paper is organized as follows: Section 2 will describe the mechatronics design of the HIT IV hand including the mechanism design and embedded control system design. Section 3 will propose the design of man-machine interface. Section 4 will give a grasping experiment to verify the effectiveness of the EMG control and sensory feedback. Finally, conclusions will be revealed in section 5.

## 2 Mechatronics Design

### 2.1 Mechanism Design

The HIT IV hand is composed of five active fingers; and each finger is actuated by a DC motor. All the motors, sensors and motion control system are integrated in the hand. The hand is about 90% of a human hand in size, weighs about 420g, whose appearance is shown in Fig.1.

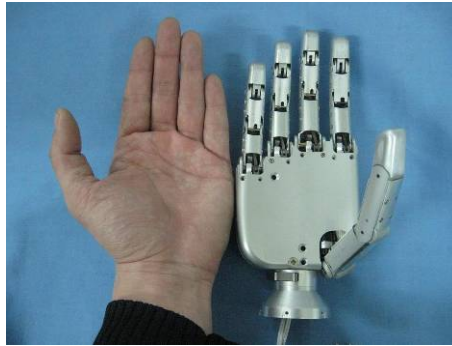


Fig. 1. Comparison between human hand and HIT IV hand



(a) spherical grasp

(b) cylindrical grasp

(c) precision grasp

Fig. 2. Grasping objects with different shapes

By taking the advantage of coupling linkage mechanism, each finger is capable to rotate around the metacarpophalangeal (MCP), proximalinterphalangeal (PIP), and distalinterphalangeal (DIP) joints for a total of 15 movable joints of the entire hand. Compared with the fixed shape fingers, not only the grasp ability is enhanced, but also the mimics motion of human hand can be achieved. Fig.2(a)(b)(c) shows the grasp ability of the hand.

### 2.2 Embedded Control System Design

The embedded control system is composed of motion control subsystem and sensory subsystem. Fig.3 shows the details of the system hardware architecture.

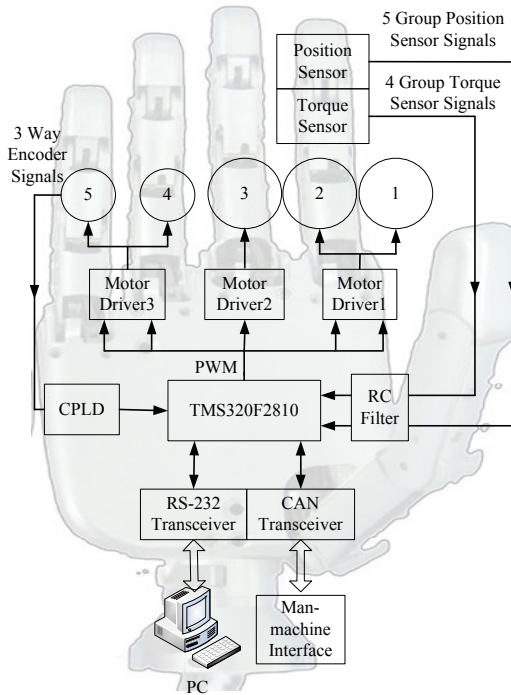


Fig. 3. Architecture of embedded control system

The sensory subsystem is crucial for both sensory feedback function and autonomous hand operation. A number of sensors, such as joint angle sensors, torque sensors and incremental magnetic encoders, are employed in the system. Position control of each finger's base joint can be obtained by means of an absolute joint angle sensor embedded in the mechanisms based on giant magneto resistance (GMR). The main advantages of the GMR sensor are their small sizes and their contactless working principle, which can avoid friction force. To protect the mechanical structure of the finger and measure the actually exerted torque of the finger, one-dimensional torque sensor is integrated into each finger's MCP joint except the little one. These torque

sensors are fundamental for impedance control and grasp force sensory feedback. Incremental magnetic encoders mounted on the motor shafts can produce feedback information of the rotation speed of the motor, which can be used in position control loop of the finger's base joint.

The motion control subsystem of prosthetic hand has to meet the requirements of small size, light weight, low power consumption, high computation speed and flexible integration (into the palm). The core of this system is a DSP chip (TMS320F2810) with 64K flash memory, 18K RAM, six independent PWM outputs, 16-channels high-speed A/D converter, two serial communication interfaces(SCI) and one controller area network (CAN) module. The A/D ports are adopted to sample the torque and the position data. The CAN module is used to realize the communication with the man-machine interface. The SCI module is used to communicate between motion control system and the upper device as PC that uses RS-232 protocol for debugging. The PWM port is to drive the H-Bridge drive chip circuitry. Three H-Bridge drive chips which incorporate internal control logic are used to drive the DC motors. A kind of CPLD chip is selected as a coprocessor for dealing with the signals of motor encoders and sending the speed and direction of motors to DSP. The HIT IV hand integrating the embedded control system PCB is shown in Fig.4.

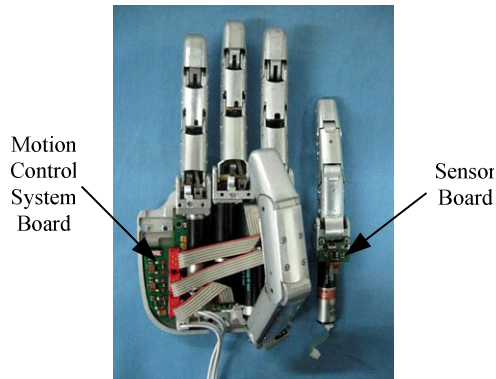


Fig. 4. Control system embedded in the HIT IV hand

### 3 Man-Machine Interface

The proposed man-machine interface is a type of interactive interface between human body and prosthetic hand for information exchanging. Its hardware is mainly composed of several EMG electrodes, a DSP board and an electrical stimulator as shown in Fig.5.

The DSP board takes charge of the main control scheme and necessary calculations, including: 1) the acquisition and processing of the myoelectric signals; 2) the generation and transmission of the action code; 3) Acquisition of the grasp force and the generation of electrical stimulus waveform data. Towards the complex operation and intensive calculations, a high performance DSP chip (TMS320F2812) is adopted as the main processor of the board, and an additional sRAM (1M) is employed to cache the real-time variables.

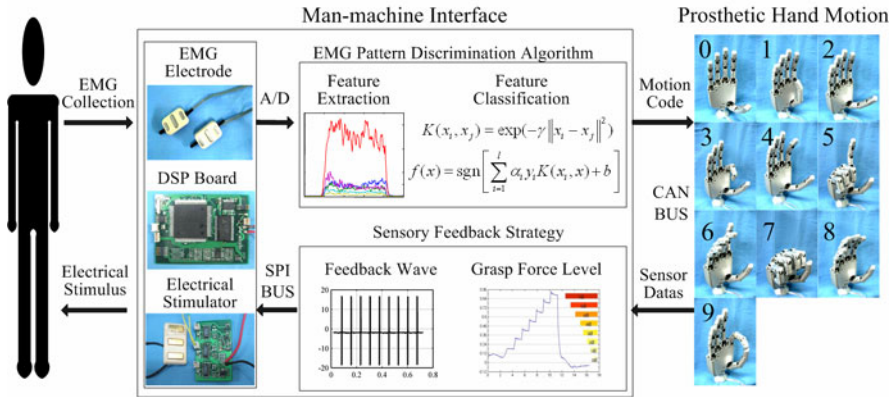


Fig. 5. Control structure of man-machine interface

Fig.5 also shows the working flow of the man-machine interface. After being amplified, filtered and rectified by the active electrodes (OttoBock 13E200=50), six channels of EMG signals are sampled at the A/D conversion ports of the DSP with a 12-bit precision. Prediction are performed on each sample using pattern recognition methods, therefore the predefined hand action modes can be recognized. Labels of the action modes are encoded and transferred to the motor controller of the prosthetic hand through CAN bus connection. According to the differences of the action modes, the prosthetic hand will perform corresponding actions automatically and acquire the force information through proprioceptive torque sensors. Based on the measurement of the grasp force, the man-machine interface translates and sends this information to the stimulator via SPI bus, meanwhile, the stimulator generates electrical stimulus wave on the surface of the human body according to the force volume.

### 3.1 EMG Processing

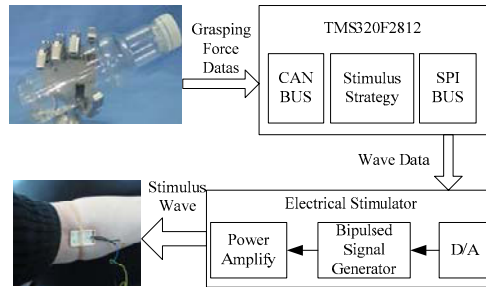
Main feature of the Otto Bock electrodes 13E200=50 is that they produce the EMG signal linear to the muscle contraction intension. Many researches [8] have revealed that this type of active electrodes can be successfully applied on the hand motion classification and hand force regression for obtaining the EMG signals. Six Otto Bock electrodes are placed on the forearm of the human body to extract the myoelectrical signals [9].The whole processing of the EMG signals includes:

- 1) Feature extraction; as stated before, because of the fine property of the output signals, sampling points at a frequency of 100Hz are directly used as the input feature vectors (6- dimension) to the classifier.
- 2) Pattern classification; a machine learning method, support vector machine (SVM) is adopted for the recognition of the action patterns based on the EMG samples. Linear kernel is used for the binary classification of any mode pairs (one-against-one) [10]. So, there will be a total of  $C_{10}^2$  classifiers, and the final decision of the mode label will be made depending on their ballots.

3) Mode encoding and transmission; the 10 modes are sequentially labelled as 0-relexion, 1-thumb flexion, 2-thumb extension, 3-index flexion, 4-index extension, 5-rest three finger flexion, 6-rest three finger extension, 7-all flexion, 8-all extension, 9-nip (thumb and index flexion). The label are directly transferred to the hand controller via CAN connection, and the controller will command the correlative motor/motors to drive the prosthetic hand's fingers to perform the current action modes according to the label, as shown in Fig.5.

### 3.2 Sensory Feedback Based on Electrical Stimulus

For getting the knowledge of the prosthetic hand's exerting force while grasping objects, thus making the amputees be able to control the force as their intentions, a sensory feedback system is proposed. The feedback system utilizes the prosthetic hand's MCP torque sensor for getting the hand's exerting force. The force data is acquired and graded by the hand's motion control system, then transmitted to the man-machine interface via CAN bus connection. The man-machine interface generates some constructive parameters for establishing the stimulation waveform based on a typical stimulation strategy, then transmits the data to the stimulator via SPI bus connection. The stimulator produces the concrete stimulus waveforms which will be applied on the human body. Fig.6 shows these operations in detail.



**Fig. 6.** The structure of the force feedback system

The feedback system utilized a self-built micro electrical stimulator, which is also shown in Fig.5. The multi-channel stimulation electrode can realize 8 stimulating grades and consumes little power. The stimulus waveforms are modulated square waves, as shown in Fig.7, where the continuous diphase square wave is utilized as the carrier wave [11],  $t_1$  is the stimulating time,  $t_2$  is the intermittent time,  $T$  is a period of the stimulating pulses. At present, 8 grades of stimulation intensity can be performed according to the scope of the grasping force (0F-1F, F denotes the maximum force), as shown in Fig.8. The pulse frequency of the stimulator can reach to 100Hz from 0Hz gradually, which makes the feeling of the electrical stimulus more and more intensive.

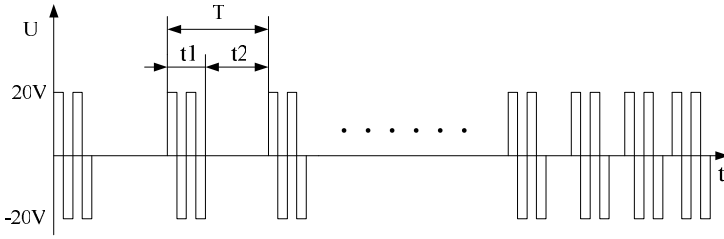


Fig. 7. Electrical stimulation waveforms

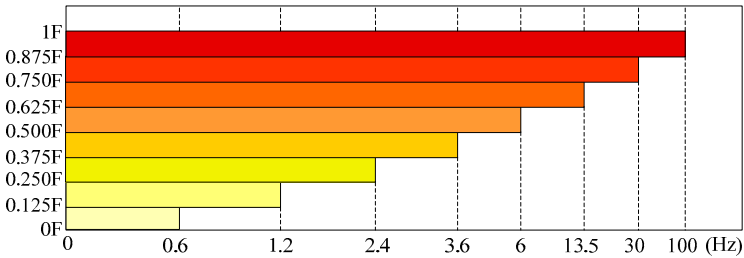


Fig. 8. Relationship between the grasping force and the grades of the stimulation

### 4 Experiment of Man-Machine Interface

In order to verify the features of man-machine interface, a grasp trial controlled by the EMG has been performed here. The hardware setup is exhibited in fig.9, which includes a prosthetic hand, a man-machine interface DSP board, an electrical stimulator, a stimulus electrode, 6 EMG electrodes, and a DC power source.

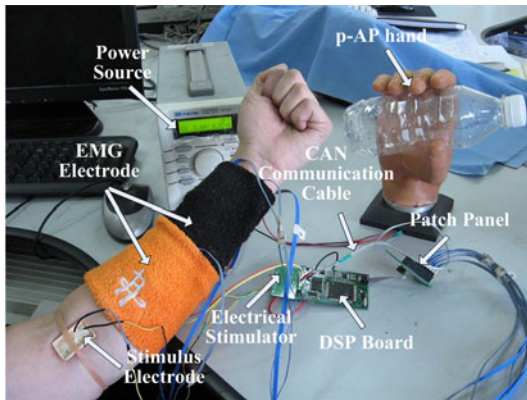
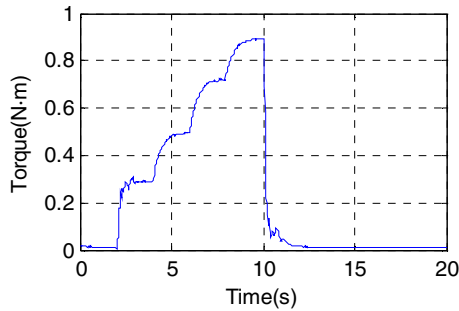


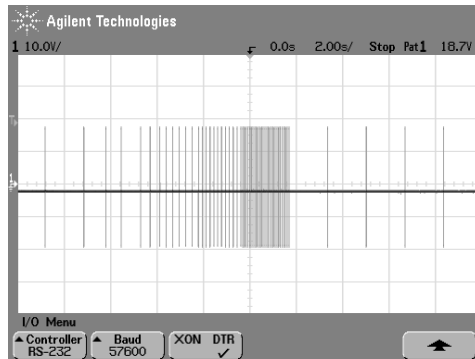
Fig. 9. Experimental setup



The experiment was performed on a healthy subject. The prosthetic hand was controlled by the subject's forearm EMG signals to hold a plastic bottle more and more tightly. Meanwhile, the information of the hand grasp force was transmitted to the subject through the graded electrical stimulation. The stimulation intensity indicates the different levels of the grasp force introduced in section 3.2. Fig.10 shows the torque curve of index finger. Several squeezing and release action can be found in Fig.10. Then, Fig.11 shows the corresponding stimulus wave, which becomes more and more serried and changes sparse at last. Through this way, the subject can get an approximate sensory feedback about the grasp force.



**Fig. 10.** Torque curve of index finger



**Fig. 11.** Stimulus wave flowing through the body surface

## 5 Conclusions

This paper developed a new highly integrated five-finger prosthetic hand with ability of interaction between human body and the prosthetic hand. Each finger of the hand was designed based on the coupling linkage principle and actuated by a DC motor individually. The embedded control system was composed of a sensory subsystem and a motion control subsystem which were integrated in the structure of prosthetic hand. A novel man-machine interface was designed for the hand based on EMG and

surface electrical stimulus, which can discriminate 10 types of hand gestures and feedback 8 levels of the grasping force intensity. Experiments were conducted using the developed system for a healthy subject. The experimental results showed that the subject can control the prosthetic hand through his EMG signals and get an approximate feeling about the grasp force at the same time.

**Acknowledgments.** This work is supported by the National High Technology Research and Development Program (Grant No. 2009AA043803), and Program for New Century Excellent Talents in University (Grant No: NCET-09-0056).

## References

1. I-Limb Hand Brochure, <http://www.touchBionics.com/document-library>
2. Cipriani, C., Zaccane, F., Stellin, G., Beccai, L., Cappiello, G., Carrozza, M.C., Dario, P.: Closed-loop Controller for a Bio-inspired Multi-fingered Underactuated Prosthesis. In: Proceedings of 2006 IEEE International Conference on Robotics and Automation, pp. 2111–2116. IEEE Press, New York (2006)
3. Carrozza, M.C., Cappiello, G., Micera, S., Edin, B.B., Beccai, L., Cipriani, C.: Design of a Cybernetic Hand for Perception and Action. *Biological Cybernetics* 95(6), 629–644 (2006)
4. Lake, C., Miguelez, J.M.: Evolution of Microprocessor Based Control Systems in Upper Extremity Prosthetics. *Technology and Disability* 2(15), 63–67 (2003)
5. Cipriani, C., Zaccane, F., Micera, S., Carrozza, M.C.: On the Shared Control of an EMG-Controlled Prosthetic Hand Analysis of User-prosthesis Interaction. *IEEE Transactions on Robotics* 24(1), 170–184 (2008)
6. Hernandez Arieta, A., Kato, R., Yokoi, H., Yu, W.W.: Development of a Multi-DOF Electromyography Prosthetic System Using the Adaptive Joint Mechanism. *Applied Bionics and Biomechanics* 3(2), 101–112 (2006)
7. Kato, R., Yokoi, H., Hernandez Arieta, A., Yu, W.W., Arai, T.: Mutual Adaptation Among Man and Machine by Using f-MRI Analysis. *Robotics and Autonomous Systems* 57(2), 161–166 (2009)
8. Castellini, C., Van Der Smagt, P.: Surface EMG in Advanced Hand Prosthetics. *Biological Cybernetics* 100, 35–47 (2009)
9. Yang, D.P., Zhao, J.D., Gu, Y.K., Wang, X.Q., Li, N., Jiang, L., Liu, H., Huang, H., Zhao, D.W.: An Anthropomorphic Robot Hand Developed Based on Underactuated Mechanism and Controlled by EMG Signals. *Journal of Bionic Engineering* 6(3), 255–263 (2009)
10. Hsu, C.W., Lin, C.J.: A Comparison of Methods for Multi-class Support Vector Machines. *IEEE Transactions on Neural Networks* 13, 415–425 (2002)
11. Scott, R.N., Caldwell, R.R., Brittain, R.H.: Sensory-feedback System Compatible with Myoelectric Control. *Medical and Biological Engineering and Computing* 18(1), 65–69 (1980)

# Dexterous and Self-adaptive Under-Actuated Humanoid Robot Hand: GCUA Hand II

Demeng Che and Wenzeng Zhang

Key Laboratory for Advanced Materials Processing Technology, Ministry of Education;  
Dept. of Mechanical Engineering, Tsinghua University, Beijing 100084, China  
demeng.che@gmail.com, wenzeng@tsinghua.edu.cn

**Abstract.** Traditional under-actuated function can make robotic hands low depend on complex control, however, it cannot ensure the hands enough flexibility to grasp or operate. A gesture-changeable under-actuated (GCUA) function is presented to make traditional under-actuated hands grasp and manipulate more dexterously, simultaneously to lower the high dependence of robotic hands on control system. This paper presents a novel under-actuated humanoid robot hand: GCUA Hand II, which is designed with GCUA function. The hand has 5 fingers and 14 DOFs. All fingers use tendon mechanisms to achieve GCUA function which includes traditional under-actuated (UA) grasping motion and special pre-bending (PB) motion. Operation principles and force analyses of the fingers are given, and the effect of GCUA function on the movements of the hand is discussed. The control system of the hand is described. The analyses and experimental results show that the GCUA function is very nice and valid. The hands with the GCUA function can meet the requirements of grasping and operating with lower control and cost, which is the middle road between traditional under-actuated hands and dexterous hands.

**Keywords:** Robot hand, under-actuated grasping, self-adaptation, pre-bending, gesture-changeable under-actuated function.

## 1 Introduction

In recent years, breathtaking developments in robotics have been witnessed all over the world. There are great numbers of robotic hands or graspers which are designed and analyzed.

In the field of dexterous hand, plentiful achievements have been made. For instance, Utah/MIT dexterous hand was designed by S. Jacobsen, et al<sup>[1]</sup>, which has four 4-DOF fingers with 32 independent tendons and 32 pneumatic cylinders<sup>[2]</sup>, the hand can be used as a high flexible tool for the study of machine dexterity; Stanford/JPL dexterous hand was designed and analyzed by J. Salisbury, et al<sup>[3]</sup>, which has three 3-DOF fingers actuated by 12 DC motors, each joint of this hand can be flexed and extended independently by one actuator; Gifu II hand, designed by H. Kawasaki, et al<sup>[4]</sup>, has 5 fingers whose all joints are actuated by servomotors, which can perform dexterous object manipulation like the human hand; Dexterous hands can achieve almost all the movements of human beings' hand, and even give some gestures which people feel

hard to pose. In reality, almost each joint of a dexterous hand needs an actuator to drive, which makes the hand high depend on control and great cost on manufacturing expenses and maintenance charges.

On the contrary, based on a very amazing feature: self-adaptation in grasping, under-actuated hands use fewer motors to drive more DOFs, which let the hands easy to manufacture and control. Many studies on under-actuated hands have been done: L. A. Dollar, et al<sup>[5]</sup>, gave a SDM robust robotic grasper which uses a single actuator to actuate 8 DOFs; Birglen, et al<sup>[6-7]</sup>, designed many kinds of under-actuated grippers and gave force analyses on them; L. Tan, et al<sup>[8]</sup>, designed a multi-fingered hand using hydraulic actuation with fluidic actuators, the hand has 14 DOFs which can bend when hydraulic pressure is applied by a water pump.

A gesture-changeable under-actuated (GCUA) function is presented to make traditional under-actuated hands grasp and manipulate more dexterously, simultaneously to lower the high dependence of robotic hands on control system. This paper presents a novel under-actuated humanoid robot hand: GCUA Hand II, which is designed with GCUA function.

## 2 Gesture-Changeable Under-Actuated Function

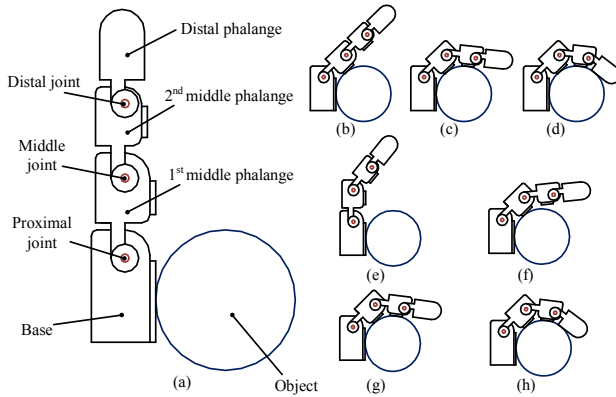
Traditional under-actuated (UA) function can drive many DOFs of a robotic hand with motors, and the number of the motors is less than that of the hand's DOFs. There are two main kinds of under-actuated hands with traditional UA function:

1) When a hand grasps objects, it cannot change its initial gesture which is fixed (almost keep straight) until its phalanges touch the objects. An example of this kind of under-actuated hand is TH-3 hand which designed by Tsinghua University<sup>[9]</sup>. TH-3 hand is based on gear-rack transmission, which cannot bend its phalanges without resistances of objects, this makes it feel hard to grasp gliding objects stably and it cannot do many gestures.

2) When a hand grasp objects, all or part of its joints must bend in a fixed order which cannot be changed when the mechanism has been designed. For instance, a multi-function mechanical hand with shape adaptation which was designed by G. Guo, X. Qian, and A. Gruver<sup>[10]</sup>, is based on linkage and gear transmission. The hand can grasp different objects with shape adaptive function. However, with its gear transmission, the rotating angles of middle joint and distal joint are proportional, which limits its dexterity.

Traditional UA function can make robotic hands low depend on sensor and control, however, it cannot ensure the hands enough flexibility to grasp or operate. A concept called gesture-changeable under-actuated (GCUA) function<sup>[11-12]</sup> is presented to make traditional under-actuated hands grasp and manipulate more dexterously, simultaneously to lower the high dependence of robotic hands on control system. The GCUA function includes two main motions: under-actuated (UA) motion and pre-bending (PB) motion, which are shown in Fig.1.

The most important feature of UA motion is self-adaptive grasping, which makes it preserve advantages of traditional under-actuated hand and lower the dependence of the hand on sensor and control system. With UA motion, the hand can grasp different objects self-adaptively.



**Fig. 1.** UA motion and PB motion of 3-joint GCUA finger. a-d) UA motion; a) Initial gesture (keep straight); b) Proximal joint rotates until 1<sup>st</sup> middle phalange touches the object; c) Middle joint rotates until 2<sup>nd</sup> middle phalange touches the object; d) Distal joint rotates until distal phalange touches the object; a-e) PB motion; e) Middle joint rotates to change the finger's initial gesture; f-h) UA motion with the changed initial gesture.

In some cases, the hand can change its initial gesture with PB motion, which can achieve different grasping postures. After getting a satisfied posture, the hand can grasp different objects with UA motion. PB motion is extremely important when high grasping dexterity and stability are required.

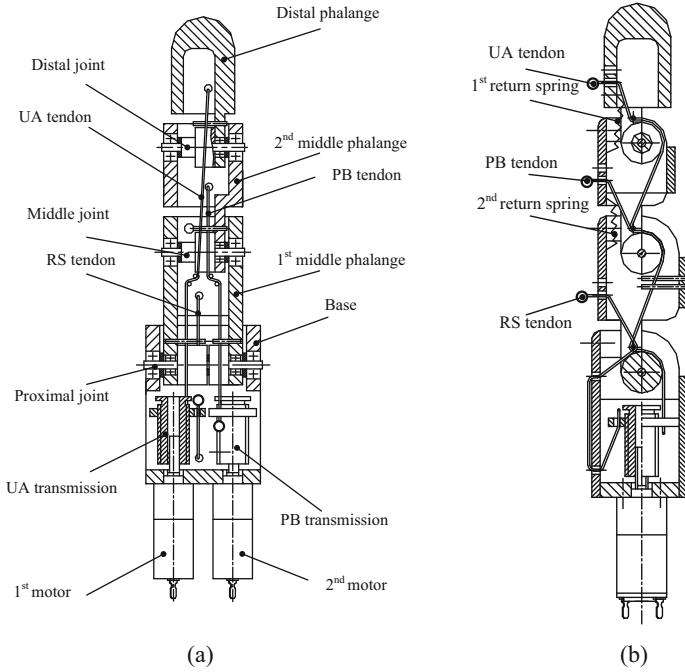
### 3 GCUA-T Finger

#### 3.1 Design of GCUA-T Finger

The GCUA-T finger with 3 joints uses tendon mechanism to achieve GCUA function, which is shown in Fig.2.

Proximal joint is fixed in base, middle joint is fixed in 1<sup>st</sup> middle phalange, and distal joint is fixed in 2<sup>nd</sup> middle phalange; UA transmission and PB transmission are both fixed in base; 1<sup>st</sup> middle phalange is sleeved with proximal joint, 2<sup>nd</sup> middle phalange is sleeved with middle joint, and distal phalange is sleeved with distal joint; 1<sup>st</sup> return spring combines 2<sup>nd</sup> middle phalange and distal phalange with its two ends, similarly 2<sup>nd</sup> return spring combines 1<sup>st</sup> middle phalange and 2<sup>nd</sup> middle phalange with its two ends.

Assembly methods of UA tendon, RS tendon, and PB tendon are shown in Fig.3. UA tendon combines UA transmission and distal phalange with its two ends, which wraps around distal joint, middle joint and proximal joint; RS tendon uses its two ends to combine 1st middle phalange and UA transmission, which wraps around proximal joint and cross through the base; PB tendon combines 2nd middle phalange and PB transmission with its two ends, which wraps around middle joint and proximal joint; The wrapping directions of UA tendon and PB tendon are the same, which are different from that of RS tendon, so that UA tendon and PB tendon will drive the finger to be flexed, on the contrary, RS tendon will drive the finger to be extended. PB tendon can also be assembled in another method, which is shown in Fig.3(d).



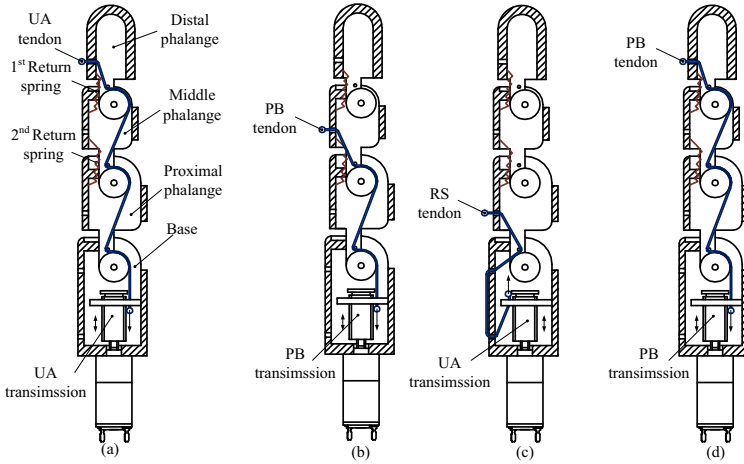
**Fig. 2.** Structure of 3-joint GCUA-T finger. a) Front cutaway view; b) Side cutaway view.

### 3.2 Work Process of GCUA-T Finger

With UA tendon and RS tendon, UA motion will proceed in this way: 1<sup>st</sup> motor rotates forward, UA transmission pulls UA tendon and RS tendon to move down. Since the wrapping directions of these two tendons are different, UA tendon will tension, simultaneously RS tendon will relax; In this case, UA tendon drives distal phalange to rotate around distal joint. With the constraint of 1<sup>st</sup> return spring and 2<sup>nd</sup> return spring, 1<sup>st</sup> middle phalange, 2<sup>nd</sup> middle phalange and distal phalange rotate around proximal joint together as a rigid body. When 1<sup>st</sup> middle phalange is blocked by objects, 2<sup>nd</sup> middle phalange and distal phalange will rotate around middle joint together, simultaneously 2<sup>nd</sup> return spring will extend; similarly, when 2<sup>nd</sup> middle phalange is blocked by objects, distal phalange will rotate around distal joint, therefore 1<sup>st</sup> return spring will extend. In this way, the finger can grasp objects self-adaptively.

When 1<sup>st</sup> motor rotates back, UA transmission will pull these two tendons up, so that UA tendon will relax, and RS tendon will tension. RS tendon will drive the finger to extend and leave the object. When UA transmission goes back to the initial status, the finger will be back to the initial gesture too.

If PB tendon is assembled in the method which is shown in Fig.3(b), PB motion will proceed like this: When 2<sup>nd</sup> motor rotates forward, PB transmission pulls PB tendon to change the finger's initial gesture, since RS tendon prevents 1<sup>st</sup> middle phalange from bending and 1<sup>st</sup> return spring makes the top two phalanges become a rigid body, 2<sup>nd</sup> middle phalange and distal phalange will rotate around middle joint together, so that the



**Fig. 3.** Assembly method of finger's tendons. a) UA tendon; b) PB tendon; c) RS tendon; d) Another assembly method of PB tendon.

finger can change its initial gesture and then grasp things with UA motion. In this process of PB motion, if either of middle phalange and distal phalange is blocked by objects, the finger will stop bending. However, when PB tendon is assembled in another method which is shown in Fig.3(d), PB motion will change. In this case, if 2<sup>nd</sup> middle phalange is blocked by objects, distal phalange can keep rotating around distal joint until it touches objects.

When 2<sup>nd</sup> motor rotates back, PB transmission will pull PB tendon up, and PB tendon will relax, after that, return springs will drive the top two phalanges to go back to their initial status.

Indeed, tendons can only restrict the movement of the finger in a single-direction, so that if one tendon relaxes, it won't have effect on other tendons, in this way, UA motion and PB motion can proceed independently. Assume  $T_{UA}$ ,  $T_{PB}$ , and  $T_{RS}$  as the driving torques of UA tendon, PB tendon and RS tendon with regard to proximal joint, and suppose that when  $T_i > 0$ , the tendon will tension, on the contrary, when  $T_i < 0$ , the tendon will relax, When  $T_{UA} > 0$ ,  $T_{RS} < 0$ , and  $T_{PB} < 0$ , both RS tendon and PB tendon relax, however, UA tendon tensions, so that UA motion proceeds and the finger is flexed; When  $T_{UA} < 0$ ,  $T_{RS} > 0$ , and  $T_{PB} > 0$ , both PB tendon and RS tendon tension, however, UA tendon relaxes, so that PB motion proceeds and the finger changes its initial gesture; When  $T_{UA} < 0$ ,  $T_{RS} > 0$ , and  $T_{PB} < 0$ , both UA tendon and PB tendon relax, on the contrary, RS tendon tensions, so that UA motion proceeds and the finger is extended.

## 4 Force Analysis of GCUA-T Finger

### 4.1 Analysis on Driving Force

Fig.4 shows the driving force when the GCUA-T finger bends or flexes. When the finger grasps objects with UA motion, the radial interaction forces between nut and

screw are  $F_a$  and  $F_{UA}$ , and the axial interaction forces between nut and screw are  $F_{t1}$  and  $F_{t2}$ . According to law of action and reaction, the equation as follows,

$$F_{t1}=F_{t2}, F_a=F_{UA} \quad (1)$$

Let the driven torque caused by 1<sup>st</sup> motor to the nut be  $T_{UA}^{M1}$  which can be expressed as,

$$T_{UA}^{M1} = \frac{1}{2} F_{t1} d_2 \quad (2)$$

where  $d_2$  is the pitch diameter of the nut.

When the motor rotates  $\theta$  the nut moves down  $S$ . Let UA transmission efficiency be  $\eta$ , according to law of the conservation of energy, the following equation can be obtained,

$$F_a S = T_{UA}^{M1} \theta \eta \quad (3)$$

Considering the geometrical relationship of UA transmission, the equation as follows,

$$P = \pi d_2 \tan \gamma \quad (4)$$

where  $\gamma$  is the lead angle of UA transmission.

Since UA transmission is single thread, according to eq.(3), the following equation can be obtained,

$$F_a (\pi d_2 \tan \gamma) = T_{UA}^{M1} \times 2\pi \eta \quad (5)$$

Simplify eq.(5), the radial interaction forces between nut and screw  $F_a$  and  $F_{UA}$  can be expressed as,

$$F_a = F_{UA} = \frac{2T_{UA}^{M1} \eta}{d_2 \tan \gamma} \quad (6)$$

According to eq.(6), when the finger relaxes with UA motion, the driven force  $F_{RS}$  which is actuated by the torque caused by 1<sup>st</sup> motor  $T_{RS}^{M1}$  can be obtained in the similar way,

$$F_{RS} = \frac{2T_{RS}^{M1} \eta}{d_2 \tan \gamma} \quad (7)$$

Similarly, when the finger pre-bends its phalange with PB motion, the driven force  $F_{PB}$  which is actuated by the torque caused by 2<sup>nd</sup> motor  $T_{PB}^{M2}$  can be expressed as follows,

$$F_{PB} = \frac{2T_{PB}^{M2} \eta_2}{d_2' \tan \gamma_2} \quad (8)$$

where  $\eta_2$  is PB transmission efficiency,  $d_2'$  is the pitch diameter of the nut as a part of PB transmission,  $\gamma_2$  is the lead angle of PB transmission.

Considering eq.(6), (7) and (8), the driving force of tendons can be adjusted, so that they can meet the requirements of grasping movements and stabilities.





**Table 1.** Specification of GCUA-T fingers

Module	Motor	Gearbox ratio	Torque (Nmm)
3-joint finger	1 <sup>st</sup> motor (UA motion)	298:1	172
	2 <sup>nd</sup> motor (PB motion)	200:1	103
Thumb	1 <sup>st</sup> motor (UA motion)	298:1	172
	2 <sup>nd</sup> motor (PB motion)	200:1	103

Let the torque of  $F_{\text{middle}}^{\text{R}}$  with regard to MP joint  $O_1$  be  $T_{\text{middle}}^{\text{R}}$  which can be expressed as:

$$T_{\text{middle}}^{\text{R}} = (l_1 \cos \theta_2 + H_2) F_{\text{middle}}^{\text{R}} \quad (10)$$

where  $H_2$  is the force arm of  $F_{\text{middle}}^{\text{R}}$  with regard to PIP joint  $O_2$ .

Similarly, let the torque of  $F_{\text{proximal}}^{\text{R}}$  with regard to MP joint  $O_1$  be  $T_{\text{proximal}}^{\text{R}}$  which can be expressed as:

$$T_{\text{proximal}}^{\text{R}} = H_1 F_{\text{proximal}}^{\text{R}} \quad (11)$$

Considering the torque and force balance of the finger, the following equation can be obtained,

$$T_{\text{M}} = T_{\text{distal}}^{\text{S}} + T_{\text{middle}}^{\text{S}} + T_{\text{distal}}^{\text{R}} + T_{\text{middle}}^{\text{R}} + T_{\text{proximal}}^{\text{R}} \quad (12)$$

When the finger grasps objects with UA motion, UA tendon will drive the finger to bend, and the driven torque  $T_{\text{M}}$  can be obtained as follows,

$$T_{\text{M}} = F_{\text{UA}} R_1 \quad (13)$$

When the finger relaxes with UA motion, RS tendon will drive the finger to flex back, and the driven torque  $T_{\text{M}}$  can be obtained as follows,

$$T_{\text{M}} = -F_{\text{RS}} R_1 \quad (14)$$

The finger sometimes needs to pre-bend to give a satisfied initial gesture before grasping, in this condition, PB tendon will drive the finger to proceed PB motion, and the driven torque  $T_{\text{M}}$  can be expressed as:

$$T_{\text{M}} = F_{\text{PB}} R_1 \quad (15)$$

The specification of GCUA-T fingers is shown in Table 1. When the finger grasp objects with UA motion, let  $l_1=28.5\text{mm}$ ,  $l_2=26.5\text{mm}$ ,  $l_3=33\text{mm}$ ,  $H_1=14\text{mm}$ ,  $H_2=13\text{mm}$ ,  $H_3=15\text{mm}$ , with the spring torque neglected, the residual torque  $T_{\text{R}}$  can be obtained as follow,

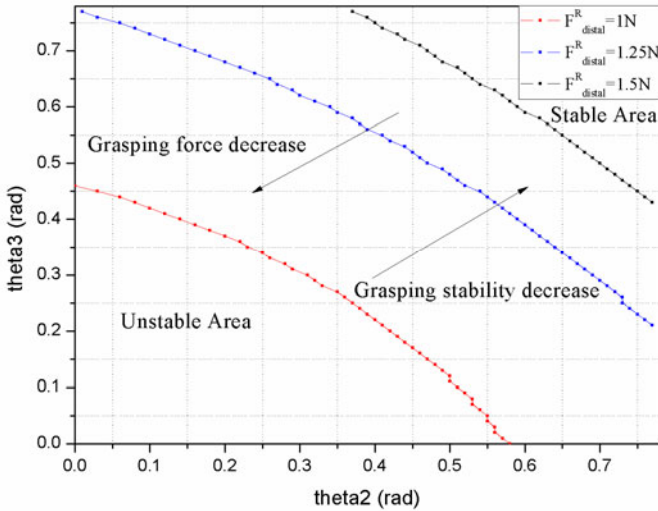
$$T_{\text{R}} = T_{\text{M}} - T_{\text{distal}}^{\text{R}} - T_{\text{middle}}^{\text{R}} - T_{\text{proximal}}^{\text{R}} \quad (16)$$

When  $T_{\text{R}}$  is not less than zero, the finger has enough power to grasp objects stably. Let  $F_{\text{middle}}^{\text{R}}=2\text{N}$ ,  $F_{\text{proximal}}^{\text{R}}=2\text{N}$ , the grasping stability area which is shown in Fig. 6, can be obtained with eq.(9), (10), (11) and (16).

As is shown in Fig.6, when the grasping angles are in the area which is below the line, the finger cannot grasp objects stably. When the grasping force of distal phalange

increases, the size of the grasping stability area decreases, and the maximum size of grasped objects decreases too. Grasping forces of middle phalange and proximal phalange have similar effects to the grasping force of distal phalange.

In fact, the magnitudes of grasping forces  $F^R_{dismal}$ ,  $F^R_{middle}$ , and  $F^R_{proximal}$  can be estimated from eq. (9), (10), (11) and (12), and the suitable motor driven torques can be determined with different demands of the magnitudes of grasping forces.



**Fig. 6.** Grasping stability area of 3-joint GCUA-T finger when UA motion proceeds. Red line:  $F^R_{distal}=1N$ ,  $F^R_{middle}=2N$ ,  $F^R_{proximal}=2N$ ; Blue line:  $F^R_{distal}=1.25N$ ,  $F^R_{middle}=2N$ ,  $F^R_{proximal}=2N$ ; Black line:  $F^R_{distal}=1.5N$ ,  $F^R_{middle}=2N$ ,  $F^R_{proximal}=2N$ ;

## 5 GCUA Hand II

Based on GCUA principle mentioned above, GCUA Hand II has 5 fingers and 14 DOFs, the thumb has 2 DOFs, and each of other fingers has 3 DOFs, which is shown in Fig.7. All of these fingers are fixed in the palm, which can work together or separately with UA motion and PB motion. The sizes of the fingers are similar with human being's: 60mm thumb, 87.5mm forefinger, 90.5mm middle finger, 84.5mm ring finger, and 75.5mm litter finger.

A Control system is designed to control all the motions of the hand, which is shown in Fig.8. This control system includes 4 parts: computer, FPGA motion controller, driver circuits, and DC motors. Computer can communicate with FPGA motion controller through USB bus, and FPGA motion controller can send PWM signals to L293D driver circuits. DC motors are connected with the output of motor driver, which can be controlled to run towards different directions with various rotate speeds.

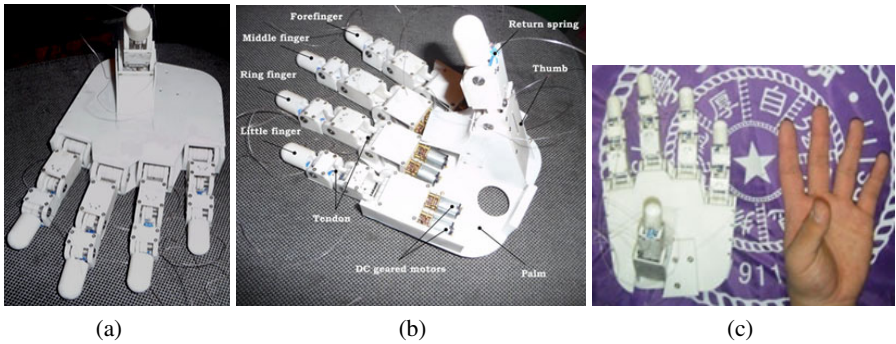


Fig. 7. GCUA Hand II

Some typical grasping actions in experiments of GCUA Hand II are shown in Fig. 9. The hand can grasp different objects self-adaptively with UA motion, simultaneously the cooperation of UA motion and PB motion enables the hand to do some simple operation like pressing buttons, holding a pen, and so on. Based on GCUA function, the hand can achieve dexterous grasping movements with low requirements for controlling it.

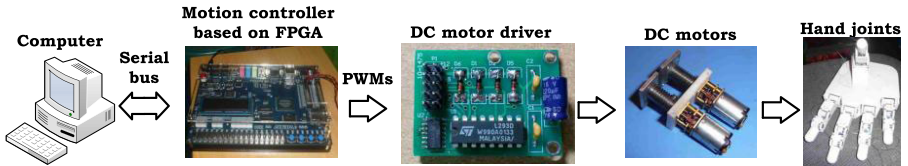


Fig. 8. Overall diagram of control system for GCUA Hand II

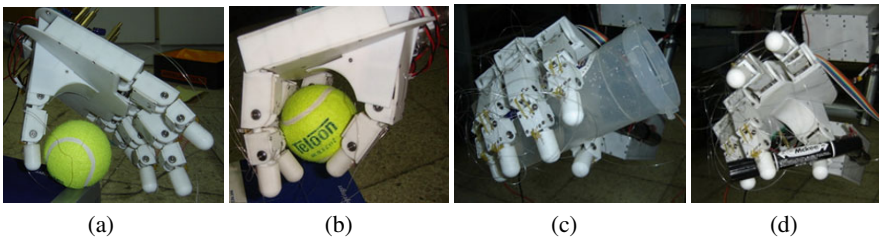


Fig. 9. Typical actions of GCUA Hand II in experiments. a, b) Grasp a ball; c) Grasp a cup; d) Manipulate a pen.

## 6 Conclusion

A novel gesture-changeable under-actuated (GCUA) function is put forward, which includes two motions: under-actuated (UA) motion and pre-bending (PB) motion. The hands with GCUA function have many advantages of traditional under-actuated hands and dexterous hand, but not many disadvantages of them.

A novel humanoid hand, called GCUA Hand II, is developed. The hand has 5 fingers and 14 DOFs, which uses tendon mechanisms to achieve GCUA function. With GCUA function, the hand can change its initial gesture with PB motion, and grasp objects self-adaptively with UA motion; simultaneously the cooperation of UA motion and PB motion enables the hand to do some humanoid operations.

The analyses and Experimental results show that the GCUA function is very nice and valid. GCUA Hand II can grasp different objects self-adaptively like traditional under-actuated hands. Furthermore, it can also operate some instruments or make humanoid poses as a dexterous hand. Although it is less dexterous than dexterous hand, it can meet many needs of grasping or operating with a lower manufacture and maintenance cost. The hands with the GCUA function will be the middle road between traditional under-actuated hands and dexterous hands.

**Acknowledgements.** This paper was funded by Hi-Tech R&D Program (2007AA04Z258), National Natural Science Foundation of China (No.50905093) and the Foundation of National Research Innovation for University Students (No.091000311).

## References

1. Jacobsen, S.C., Iversen, E.K., Knutti, D.F., et al.: Design of the UTAH/M.I.T. dextrous hand. In: Proc. IEEE Inter. Conf. on Robotics and Automation, pp. 1520–1532. IEEE Computer Society Press, San Francisco (April 1986)
2. Guo, G., Gruver, W.A., Qian, X.: A new design for a dexterous robotic hand mechanism. IEEE Control Systems Magazine 12(4), 35–38 (1992)
3. Salisbury, J.K., Craig, J.J.: Articulated hands: force control and kinematic issues. Inter. J. of Robotics Research 1(4), 4–17 (1982)
4. Kawasaki, H., Komatsu, T., Uchiyama, K., et al.: Dexterous anthropomorphic robot hand with distributed tactile sensor: Gifu hand II. In: Proc. IEEE Inter. Conf. on Systems, Man and Cybernetics, Tokyo, Japan, pp. 782–787 (October 1999)
5. Dollar, A.M., Howe, R.D.: The SDM hand as a prosthetic terminal device: a feasibility study. In: Proc. the 2007 IEEE 10th International Conference on Rehabilitation Robotics, Noordwijk, The Netherlands, pp. 978–983 (June 2007)
6. Birglen, L., Gosselin, C.M.: Kinetostatic analysis of underactuated fingers. IEEE Trans. on Robotics and Automation 20(2), 211–221 (2004)
7. Birglen, L., Gosselin, C.M.: Fuzzy enhanced control of an underactuated finger using tactile and position sensors. In: Proc. IEEE Inter. Conf. on Robotics and Automation, Barcelona, Spain, pp. 2320–2325 (April 2005)
8. Tan, L.Q., Xie, S.Q., Lin, I.C., et al.: Development of a multifingered robotic hand. In: Proc. IEEE Inter. Conf. on Information and Automation, Zhuhai/Macau, China, pp. 1541–1545 (June 2009)
9. Zhang, W., Che, D., Liu, H., et al.: Super under-actuated multi-fingered mechanical hand with modular self-adaptive gear-rack mechanism. Industrial Robot: An International Journal 36(3), 255–262 (2009)
10. Guo, G., Qian, X., Gruver, A.: Multi-function Mechanical Hand with Shape Adaptation: U.S. Patent: No. 5378033[P] (January 1995)

11. Zhang, W., Che, D., Chen, Q., et al.: A dexterous and self-adaptive humanoid robot hand: gesture-changeable under-actuated hand. In: Proc. 2nd Inter. Conf. on Intelligent Robotics and Applications, Singapore, pp. 515–525 (December 2009)
12. Zhang, W., Che, D., Chen, Q., et al.: Development of gesture-changeable under-actuated humanoid robotic finger. *Chinese Journal of Mechanical Engineering* 23(2), 142–148 (2010)

# A Novel Coupled and Self-adaptive Under-actuated Grasping Mode and the COSA-DTS Hand

Jie Sun, Wenzeng Zhang, and Haitao Sun

Key Laboratory for Advanced Materials Processing Technology, Ministry of Education;  
Dept. of Mechanical Engineering, Tsinghua University, Beijing 100084, China  
sunjie\_2004@sina.com, wenzeng@tsinghua.edu.cn

**Abstract.** A novel category of humanoid robotic hands: COSA (Coupled and Self-Adaptive) hand is introduced in this paper. If a hand is called the COSA hand, which means that the fingers of the hand grasps objects with a special COSA grasping mode, which includes two processes, first is a coupled grasping process, then the second is a self-adaptive grasping process. The COSA grasping mode can take many advantages. A new COSA finger unit with double-tendon mechanism is designed based on the COSA function. The COSA finger is a new category of robotic finger which is different from conventional kinds of robotic fingers, such as coupled fingers or self-adaptive fingers. The shortages of traditional coupled fingers or self-adaptive fingers are avoided by the new COSA fingers. Based on the two joints finger, a three joints finger is designed and the COSA-DTS Hand with double-tendon-slider mechanism is designed based on it. In this paper, the core opinions about COSA fingers and COSA hands are presented, discussed and analyzed in detail via the force and motion simulations. The hand has 5 fingers, 15 joints and 6 motors. All fingers of the hand are COSA-DTS fingers. The hand is more similar to human hand in appearance and actions, able to more dexterously and stably grasp different objects than tradition coupled or self-adaptive under-actuated hands.

**Keywords:** Multi-fingered hands, Under-actuated finger, Coupled finger, Self-adaptive finger, Grasping mechanism.

## 1 Introduction

If a man-made mechanical device is called robotic hand or gripper, the device should has a base which is “palm” and multiple linkages connected to the palm which are “fingers”. One of the key technologies to design a robotic hand or gripper is how to design the mechanism of the fingers aiming to bend the finger around different objects grasped.

There are two kinds of robotic fingers as two poles: rigid fingers and active fingers.

a) Rigid fingers, noted as the number “1”.

In the fingers there is none joint or there are several pseudo rigid joints. The fingers are often used in industrial grippers or clamps, which are not robotic finger in some extent. Plenty of clamps have used in industry. The advantage of the rigid finger is that

the finger has fewest actuators so that the cost of the finger is low, and the finger is easy to control. The disadvantage of the rigid finger is that the fingers are not able to self-adapt objects grasped.

b) Active fingers, noted as the number “2”.

In the fingers there are multiple joints and all joints are active driven by actuators. The robotic fingers are called dexterous fingers. For example, Utah/MIT Hand<sup>[1, 2]</sup>, Karlsruhe Hand<sup>[3]</sup>, Robonaut Hand<sup>[4]</sup>, DLR-II Hand<sup>[5]</sup>, Shadow V5 Hand<sup>[6]</sup>, Japanese high speed dexterous hand<sup>[7]</sup> and domestic BH series hands by BUAA (Beijing University of Aeronautics and Astronautics), HIT series hands by HIT (Harbin Institute of Technology).

The advantages of the dexterous fingers are listed here: they are quite agile, they have more joints, are able to adapt different objects grasped like human fingers. The disadvantages of the dexterous fingers are found here: more actuators in one finger make its control very difficult even if the finger only grasps a cup of water, the cost of dexterous fingers are very expensive.

## 2 The Under-Actuated Finger

Due to the disadvantages of two kinds of fingers mentioned above, a special fingers were proposed several decades ago, which were hoped to have both advantages of two kinds of fingers mentioned above, and have none of disadvantages of either rigid fingers or dexterous fingers. The special fingers are called under-actuated fingers. The word “under-actuated” or “under-actuation” is written as UA for short in this paper.

How to produce UA fingers? The number of actuators in one UA finger is less than the degrees of freedom (DOFs) of the finger. The UA finger is a special kind of fingers between the number “1” of rigid fingers and the number “2” of dexterous fingers. Therefore, the number from “1” to “2” stands for the UA fingers.

The mechanism principle of UA finger is to make the several joints rotate according to some order driven by one motor or other actuator.

For example, there is a finger with several joints: the lower joint which is in the root of the finger and close to palm, the middle joint and the upper joint which is in the end of the finger and far from the palm. There is only one motor embedded in the root of the finger. There is a mechanism from the motor to several joints. There are three kinds of UA fingers according to the motion order of joints in the fingers.

1) Joints rotate from end to root. This case can be written as the number “1.3” and the finger is called inverse order UA finger. The grasping process of inverse order UA finger is shown in Fig. 1.

2) Joints simultaneously rotate. This case can be written as “1.5” and the finger is called coupled UA finger or coupled finger.

For example, a coupled finger with 2 joints. The grasping process of coupled UA finger is shown in Fig. 2. If the proximal joint revolves, the distal joint will revolve by an angle with a fixed proportion of the proximal joint. A fixed rotating relationship between its joints is designed to reduce the number of actuators. The ratio of rotation angles between coupled phalanges is set to be fixed during mechanical design.

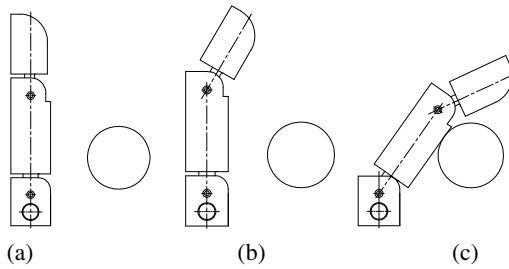
For example, TBM Hand<sup>[8]</sup>, MANUS-Hand<sup>[9]</sup>, Southampton Hand<sup>[10]</sup> and iCub Hand<sup>[11]</sup> have many coupled UA fingers.



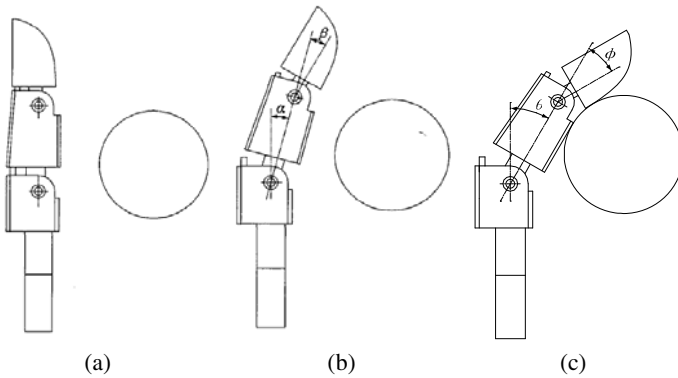
The advantage of coupled UA finger: it is similar to human finger in grasping actions, has fewest motors and its control is simple. The shortage of coupled finger: it cannot self-adapt different objects.

3) Joints rotate from root to end. This case can be written as “1.7” and the finger is called self-adaptive UA finger, passive adaptive UA hand or even often UA Finger in short. The grasping process of self-adaptive UA finger is shown in Fig. 3.

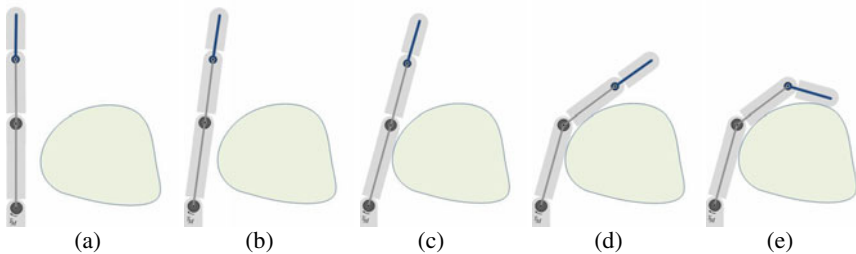
The inverse order UA fingers can be omitted since they are only used in some special situations. Traditional under-actuated (UA) finger mechanisms are divided into two main kinds: coupled UA finger and self-adaptive UA finger. General under-actuated fingers have two types: coupled UA finger and self-adaptive UA finger.



**Fig. 1.** The inverse under-actuated grasping process of 2 joints UA finger



**Fig. 2.** The coupled grasping process of 2 joints UA finger



**Fig. 3.** The under-actuated grasping process of 3 joints UA finger

The grasping adaptation to shapes and sizes of objects grasped is the main function. The fingers are able to self-adaptively grasp different objects of different shapes and sizes. The fingers can grasp a big range of sizes and shapes of objects and the grasp is stabled.

For example, there are many hands which have self-adaptive UA fingers: SARAH Hand<sup>[12, 13]</sup>, LARM Hand<sup>[14]</sup>, the under-actuated hand by HIT<sup>[15, 16]</sup>, the under-actuated hand by BH Univ.<sup>[17]</sup> and TH-3R Hand<sup>[18, 19, 20]</sup>.

Although the capability of the self-adaptive UA finger has been improved relative to the coupled UA finger, the self-adaptive UA finger has still some serious disadvantages: the fingers are not able to rotate its middle joints before its touch the grasped object, therefore the grasping capability of the finger is limited and the grasping actions of small objects are far from the actions of a human finger.

To combine the advantages of two kinds of fingers for improving the capability, a novel grasping mode: parallel coupled and self-adaptive (COSA) grasping mode is presented in this paper. A novel finger based on the COSA grasping mode is designed. to reach improved grasping capability. A novel humanoid robot hand based on this new technology is designed in this paper.

### 3 The Coupled and Self-adaptive Grasping Mode

A novel grasping mode: Coupled and Self-adaptive (COSA in short) grasping mode is presented. The action process of the COSA finger is divided into two stages.

a) The first stage: coupling grasping stage. When an actuator in base drives the proximal joint rotate by an angle  $\theta$ , the distal joint rotates coupling with the proximal joint by a same angle  $\theta$ . The proportion between two joints is 1:1. This process will continue until one of the phalanges touches the grasped object.

b) The second stage: self-adaptive grasping stage.

With the rotating of the finger, the proximal phalanx rotates by the angle  $\theta$  and touches the grasped object, but the distal phalanx has not yet touched the object, the grasping will continue.

A movable slider as a receiver is set in the slot of the proximal phalanx. The slider can move freely in the slot. The moving range of the slider is small, such as 4-6 mm. The slider protrudes from the surface of the proximal phalanx. The slider is very important part. The coupled mechanism is connected by the slider from the proximal joint to distal joint.

At the same time, the slider is a part of the self-adaptive mechanism. Once the object touches the slider, the slider will glide towards the inner of finger so that the coupled mechanism is broken and the self-adaptive mechanism is activated.

At this time, the proximal phalanx turns by a small angle  $\Delta\theta$ , the slider pans a small distance. Due to the self-adaptive mechanism from slider to the distal phalanx, the distal phalanx will rotate by a bigger angle  $\alpha$  until it touches the object. In the second stage when the proximal phalanx (in fact, the slider) touches the object, the coupled transmission mechanism is replaced by the self-adaptive transmission mechanism automatically. This stage is a self-adaptive grasping process.

## 4 The COSA-DTS Finger

### 4.1 The 2-Joint COSA-DTS Finger

A 2-joint COSA finger is designed for realizing coupled and self-adaptive grasping mode. The principle of the finger is shown in Fig. 4. In the Figure, the finger mainly consists of a DC motor mounted in a base 1, a proximal phalanx 3, a lower tendon 4, a lower pulley 2, a lower slider 5, an upper slider 6, an upper tendon 7, an upper pulley 8, a distal phalanx 9, a compression spring 10. The special mechanical design is shown in Fig 5.

The proximal joint-shaft, on which the proximal phalanx is mounted, is sleeved in the base. The distal joint-shaft is sleeved in the proximal phalanx. The distal phalanx is mounted on the distal joint-shaft and fixed with it. The lower pulley is sleeved on the proximal joint-shaft and the lower pulley is fixed with the base. The lower tendon meshes with the lower pulley. The lower slider and the upper slider are both set in two slots of the proximal phalanx. The moving directions of the two sliders are parallel. A surface plane of the lower slider touches a surface plane of the upper slider when the finger is at the initial straight status so that the lower slider is able to push the upper one towards the inner of the proximal phalanx. The lower tendon penetrates the lower slider, and the upper tendon penetrates the upper slider. The upper pulley, meshing the upper tendon, is sleeved on the distal joint-shaft and fixed with the latter. The tension spring connects the lower slider to the proximal phalanx. The torsion spring is sleeved on the distal joint-shaft and two side of the spring connects the distal joint-shaft to the distal phalanx.

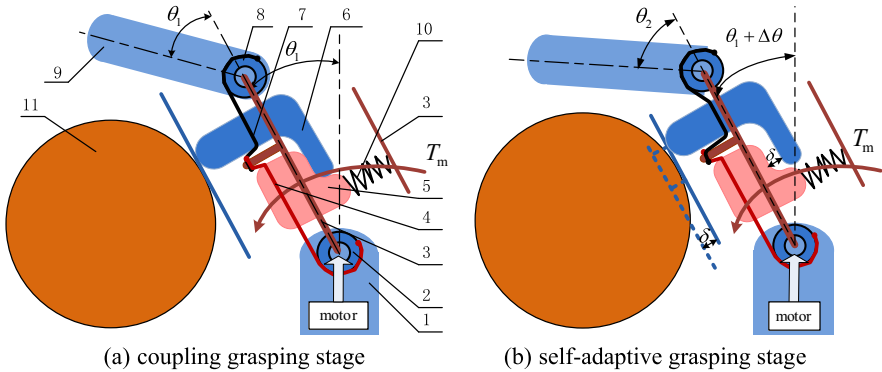
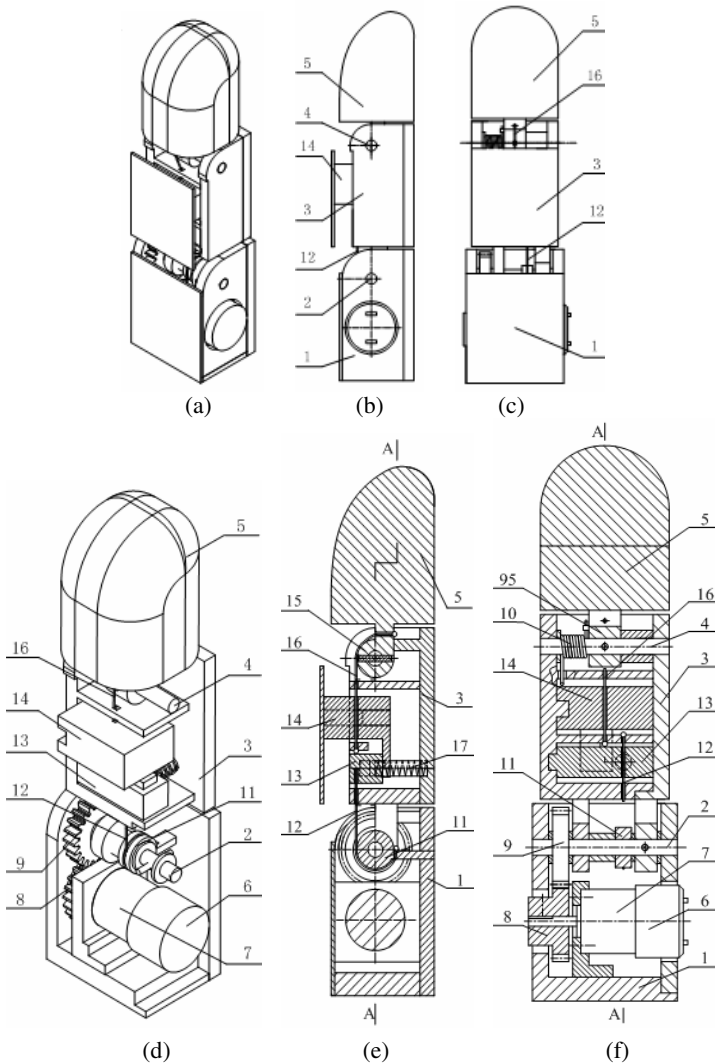


Fig. 4. The principle of the COSA-GRS finger unit

### 4.2 Grasping Principle

The grasping process of 2-joint COSA-DTS finger is shown in Fig. 6. The initial configuration of the finger is that the finger keeps straight with respect to the base shown in Fig 6a. The distal phalanx is parallel to the proximal phalanx due to the effect of the torsion spring.

The finger is provided with two grasping stage, detailed as followed, respectively. The first stage is coupling grasping motion. The motor torque is transmitted via the

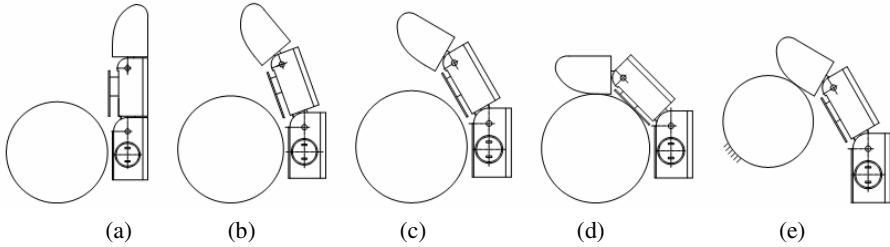


**Fig. 5.** Mechanical design of 2 joints COSA-DTS finger. 1-base; 2-proximal joint-shaft; 3-proximal phalanx; 4-distal joint-shaft; 5-distal phalanx; 6, 7-DC gear motor; 8, 9-gear; 10-torsion spring (adjusting grasping force); 11-lower pulley; 12-lower tendon; 13-lower slider; 14-upper slider; 15-upper pulley; 16-upper tendon; 17-tension spring.

reducer and the gear transmission mechanism to the proximal joint-shaft. The proximal phalanx is driven to rotate the proximal joint-shaft. The finger bends towards the object. Due to the rotation of the proximal phalanx and the lower pulley is fixed with the base, the lower tendon moves simultaneously towards the inner of the proximal phalanx. As a result the lower slider moves together with the lower tendon and pushes the upper slider. The upper tendon moves with the upper slider so that the upper pulley rotates, in the same direction with the proximal phalanx. The distal phalanx rotates with

respect to the proximal phalanx until “the proximal phalanx” (actually the upper slider) contacts the object (Fig. 6a-b-c).

It should be noted that the radius of the lower pulley and the upper pulley are the same. This leads to the ratio of the rotating angles of two phalanges is 1:1.



**Fig. 6.** Grasping process of 2 joints COSA-DTS finger

The second stage is self-adaptive grasping motion starts at the end of the first stage shown in Fig. 6c. The upper slider is pushed into the finger by the grasped object when “the proximal phalanx” (actually the upper slider) is blocked by the object. The proximal phalanx continues to rotate a small angle, which leads to the lower tendon with the lower slider pans a small distance because the diameter of the lower pulley is small (such as 5 mm). Due to the distance from the touching point between the object and the upper slider is larger, the same rotating angle of the proximal phalanx make that the moving distance of the upper slider with the upper pulley is larger than the lower slider with the lower tendon. In other words, the upper slider with the upper tendon moves more quickly than the lower slider with the lower tendon. So the upper slider does not contact with the lower slider any longer, which realizes natural decoupling of two coupled joints. The quickly moving upper tendon with the upper slider leads to that the distal phalanx fixed with the upper pulley rotates more quickly than the proximal phalanx. The second stage continues until the distal phalanx touches the object, at this time, all of phalanges have touched the grasped object. The grasping process is successfully over (Fig. 6c-d). Whatever different shapes and sizes of the objects, the grasping process is always effective.

In the second stage, the rotating angle of the distal phalanx is not proportional to the rotating angle of the proximal phalanx any longer.

If in the first stage, the distal phalanx touches the grasped object earlier than the proximal phalanx (Fig 6e), grasping process is over, which realizes a stable pinch.

The coupling grasping and self-adaptive grasping are both realized in the finger. The motion process of opening the finger has the same principle with the grasping motion process addressed above.

### 4.3 The 3-Joint COSA-DTS Finger

In fact, one can get multiple-joint COSA-DTS finger just using the same mechanism principle of the 2-joint finger. Several previous unit mechanisms can be arranged in series to produce a super COSA (high coupled and high self-adaptive) finger. The 2-joint COSA-DTS finger can be considered as a unit for constructing COSA-DTS

fingers with different number of joints. This feature can be utilized to design the index, middle, ring and little fingers of a robot hand.

Using the same mechanism of the 2-joint finger, a 3-joint COSA-DTS finger is designed, shown in Fig. 7. There are two set of the double-tendon-slider mechanisms in the 3-joint finger.

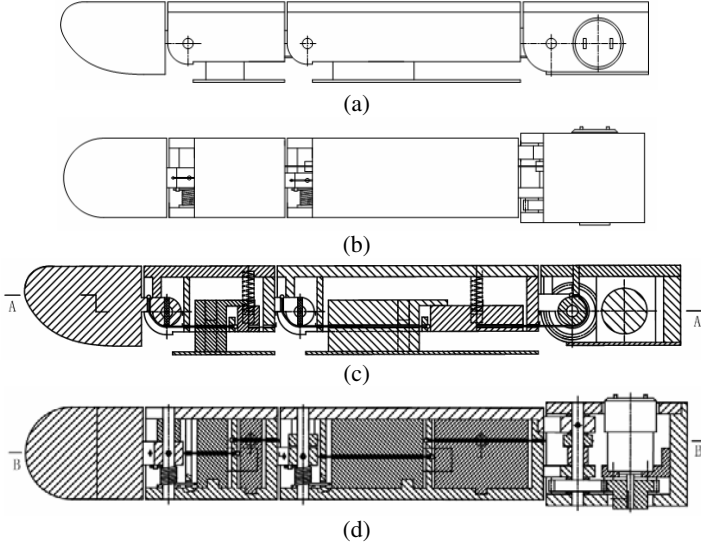


Fig. 7. Mechanical design of 3 joints COSA-DTS finger

## 5 Force Analysis

The forces of the 2-joint COSA finger when the finger grasps an object are shown in Fig. 10.

- Where,  $f_1$ : the force brought to the object by the proximal phalanx, N
- $f_2$ : the force brought to the object by the distal phalanx, N
- $f_s$ : the force brought by the tension spring, N
- $f_m$ : the force exerted by the upper slider towards the upper tendon, N
- $f'_m$ : the anti-force of the  $f_m$ , N
- $r_2$ : the radius of the upper pulley, mm
- $l_1$ : the length of the proximal phalanx, mm
- $T_m$ : the torque of the motor, Nm
- $h_1$ : the force arm of  $f_1$  with respect to the center of the lower pulley, mm
- $h_2$ : the force arm of  $f_2$  with respect to the center of the upper pulley, mm
- $s$ : the distance compressed of the tension spring, mm
- $\Delta s$ : the move distance of the lower slider, mm
- $k$ : the stiffness coefficient of the tension spring, N/mm
- $\theta_1$ : the rotation angle of the proximal phalanx relative to the base,  $^\circ$
- $\theta_2$ : the rotation angle of the distal phalanx relative to the proximal phalanx,  $^\circ$

The analysis here is based on a static model which the friction is also ignored. A set of formulas can be obtained according to the mechanism principle of double-tendon transmission mechanism and mechanics, transmission relation and geometric relation:

$$\begin{cases} T_m = f_1 h_1 + f_2 (h_2 + l_1 \cos \theta_1) \\ f_2 h_2 = f_m r_2 \\ f'_m = f_1 + f_s \\ f_s = k \Delta s \\ \Delta s = (\theta_1 + \theta_2) r_2 \end{cases} \quad (5)$$

According to equation (5), one gets the below results:

$$\begin{cases} f_1 = \frac{T_m h_2 - k(\theta_1 + \theta_2) h_1 h_2 r_2}{h_1 h_2 + r_2 (h_2 + l_1 \cos \theta_2)} + k(\theta_1 + \theta_2) r_2 \\ f_2 = \frac{T_m r_2 - k(\theta_1 + \theta_2) h_1 r_2^2}{h_1 h_2 + r_2 (h_2 + l_1 \cos \theta_2)} \end{cases} \quad (6)$$

When  $T_m=5000$  mNm,  $r_2=3$  mm, the range of  $\theta_1$  and  $\theta_2$  from  $0\sim 90^\circ$ ,  $h_1=20$  mm,  $h_2=10$  mm,  $l_1=40$  mm,  $k=0.5$  N/mm, then the relationships of  $f_1, f_2, \theta_1$  and  $\theta_2$  can be obtained. The relationship of  $f_1, \theta_1$  and  $\theta_2$  is shown in Fig. 11.

The conclusions from the figures are listed below.

- 1) If  $\theta_1$  increases, then the  $f_1$  and  $f_2$  will change slightly.
- 2) If  $\theta_2$  increases, then the  $f_1$  and  $f_2$  will increase distinctly.
- 3) Whatever the  $\theta_1$  and  $\theta_2$  are any values, the  $f_1$  and  $f_2$  will always keep positive and the amount of  $f_1$  and  $f_2$  are so large that the finger can be effective to grasp objects stably.

These simulation results show that the COSA-DTS finger is effective: it can stably grasp objects.

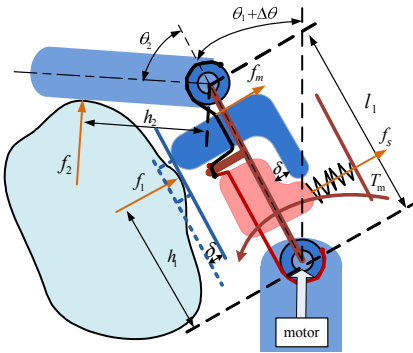


Fig. 10. Force analysis of the COSA-DTS finger

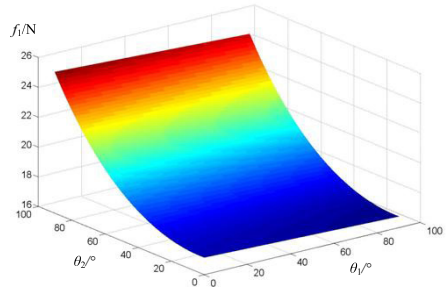


Fig. 11.  $f_1$  as a function of  $\theta_1$  and  $\theta_2$

## 6 A Humanoid Robotic Hand: The COSA-DTS Hand

A humanoid robot hand with five COSA-DTS fingers is designed, called COSA-DTS Hand. The COSA-DTS Hand has 5 fingers (all COSA-DTS fingers), 15 joints (9 COSA-DTS joints) and only 6 motors, similar to human hand in appearance and actions. The COSA-DTS Hand can do more gestures and more stably grasp objects than traditional coupled hands or self-adaptive hands.

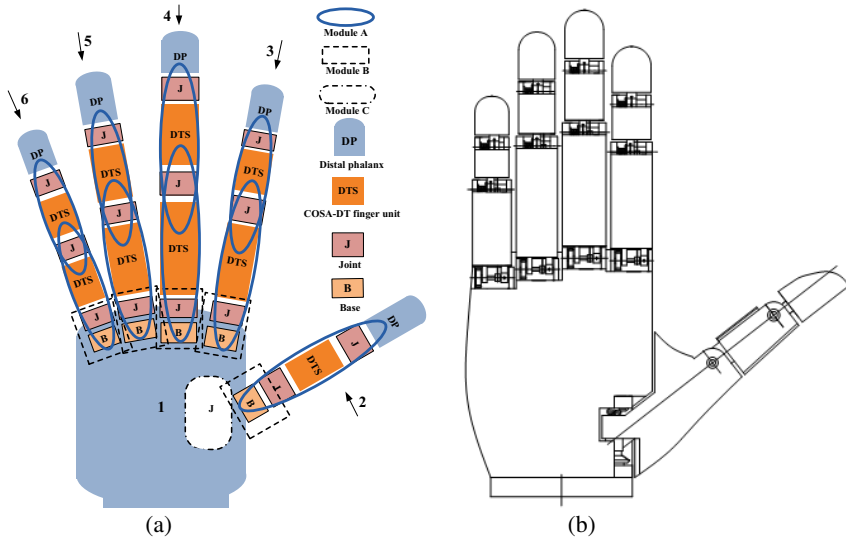
The joints of the hand can be divided into three classes of modules (Fig. 12a): Module A, B and C. The appearance of COSA-DTS Hand is shown in Fig. 12b, Module A is the COSA-DTS under-actuated finger unit above-mentioned. Module B is the active foot joint of five fingers with one embedded motor, reducer and gear transmission. Module C is the swing joint of the thumb in which there is one embedded motor.

COSA-DTS Hand is the same size with an adult's hand, with a length 200mm, palm length of 96mm, palm width of 76mm, palm thickness of 30mm with rotation angle range of 0~90° for all joints.

All the fingers in COSA-DTS Hand are designed using modularized structure. The 4 fingers except the thumb have the same structure with different sizes, with 1 motor driving 3 joints. The palm has 1 motor driving the side swinging of the base of the thumb, while the thumb has 1 motor driving 2 joints.

COSA-DTS Hand is light weight and low cost. Its motors, the transmission mechanisms and controlling circuits are completely hidden inside the hand. It can grasp objects with changeable forces, coupling grasping and self-adaptive to the shape of the grasped objects.

Compared to TH-1 Hand, TH-2 Hand, and TH-3B Hand designed and developed by Tsinghua University, COSA-DTS Hand has not only 5 fingers controlled separately,



**Fig. 12.** The COSA-DTS Hand. B-base; J-joint; 1-palm; 2-thumb; 3-index; 4-middle finger; 5-ring finger; 6-little finger



more joint DOFs, more self-adaptive to the shapes of the objects (such as sphere), but also the ability to rotate multiple joints in each finger before contacting objects with better grasping performance.

## 7 Conclusion

A novel category of humanoid robotic hands: COSA (Coupled and Self-Adaptive) hand is introduced in this paper. If a hand is called the COSA hand, which means that the fingers of the hand grasp objects with a special COSA grasping mode, which includes two processes, first is a coupled grasping process, then the second is a self-adaptive grasping process.

A new COSA finger unit with double-tendon mechanism is designed based on the COSA function. The COSA finger is a new category of robotic finger which is different from conventional kinds of robotic fingers, such as coupled fingers or self-adaptive fingers. The shortages of traditional coupled fingers or self-adaptive fingers are avoided by the new COSA fingers.

Based on the two joints finger, a three joints finger is designed and the COSA-DTS Hand with double-tendon-slider mechanism is designed. The hand has 5 fingers, 15 joints and 6 motors. All fingers of the hand are COSA-DTS fingers. The hand is more similar to human hand in appearance and actions, able to more dexterously and stably grasp different objects than tradition coupled or self-adaptive under-actuated hands.

## Acknowledgements

This research was funded by Hi-Tech R&D Program of China (No. 2007AA04Z258) and National Natural Science Foundation of China (No. 50905093).

## References

1. Jacobsen, S.C., Wood, J.E., Knutti, D.F., et al.: The Utah-MIT dextrous hand: work in progress. *The Int. Journal of Robotics Research* 3(4), 21–50 (1984)
2. Jacobsen, S., Iversen, E., Knutti, D., et al.: Design of the Utah/MIT Dexterous Hand. In: *Proc. of IEEE Inter. Conf. on Robotics and Automation*, pp. 1520–1532. IEEE Computer Society Press, San Francisco (April 1986)
3. Wöhlke, G.: A programming and simulation environment for the Karlsruhe dextrous hand. *Journal of Robotics and Autonomous systems* 9, 243–263 (1990)
4. Lovchik, C., Aldridge, H., Diftler, M.: Design of the NASA Robonaut Hand. In: *Proc. of ASME Dynamics and Control Division, DSC*, vol. 67, pp. 813–830. American Society of Mechanical Engineers, New York (1999)
5. Butterfass, J., Grebenstein, M., Liu, H., et al.: DLR-Hand II: Next generation of a dextrous robot hand. In: *Proc. of IEEE Inter. Conf. on Robotics and Automation*, pp. 109–104 (2001)
6. Reichel, M.: Transformation of shadow dextrous hand and shadow finger test unit from prototype to product for intelligent manipulation and grasping. In: *Intelligent Manipulation and Grasping, Inter Conf.*, Genova, Italy, July 1-2 (2004)

7. Chua, P.Y., Bezdicek, M., Davis, S., Gray, J.O., et al.: Tele-operated high speed anthropomorphic dextrous hands with object shape and texture identification. In: IEEE IROS, Beijing, China, pp. 4018–4023 (October 2006)
8. Dechev, N., Cleghorn, W., Naumann, S.: Multiple finger, passive adaptive grasp prosthetic hand. *Mechanism and Machine Theory* 36, 1157–1173 (2001)
9. Dubey, V.N., Crowder, R.M.: Grasping and control issues in adaptive end effectors. In: Proc. of ASME on Design Engineering Technical Conf. and Computers and Information in Engineering Conf., pp. 1–9 (2004)
10. Pons, J.L., Rocon, E., Ceres, R., et al.: The MANUS-HAND dexterous robotics upper limb prosthesis: mechanical and manipulation aspects. *Autonomous Robots* 16, 143–163 (2004)
11. Davis, S., Tsagarakis, N.G., Caldwell, D.G.: The Initial Design and Manufacturing Process of a Low Cost Hand for the Robot iCub. In: IEEE Humanoids, Daejeon, Korea, pp. 40–45 (2008)
12. Laliberte, T., Gosselin, C.: Simulation and Design of Under-actuated Mechanical Hands. *Mechanism and Machine Theory* 33(1/2), 39–57 (1998)
13. Birglen, L., Gosselin, C.: On the force capability of underactuated fingers. In: Proc. of IEEE Inter. Conf. on Robotics and Automation, Taipei, pp. 1139–114 (April 2003)
14. Wu, L., Ceccarelli, M.: A Numerical Simulation for Design and Operation of an Under-actuated Finger Mechanism for LARM Hand. *Mechanics Based Design of Structure and Machines* 37(1), 86–112 (2009)
15. Yang, D., Zhao, J., Gu, Y., et al.: An Anthropomorphic Robot Hand Developed Based on Underactuated Mechanism and Controlled by ENG Signals. *J. of Bionic Engineering* 6, 255–263 (2009)
16. Liu, H., Gao, X., Shi, S.: Under-actuated self-adaptive artificial hand. Chinese Patent: CN 1292719C (2007)
17. Guo, W., Li, J., Su, W., et al.: Designing research of 3DOF humanoid hand with self - adaptation grasping. *Journal of Machine Design* 21(9), 17–19 (2004) (in Chinese)
18. Zhang, W., Che, D., Liu, H., et al.: Super Under-actuated Multi-fingered Mechanical Hand with Modular Self-adaptive Gear-rack Mechanism. *Industrial Robot: An Int. J.* 36(3), 255–262 (2009)
19. Zhang, W., Che, D., Chen, Q., et al.: A dexterous and self-adaptive humanoid robot hand: gesture-changeable under-actuated hand. In: Proc. 2nd Inter. Conf. on Intelligent Robotics and Applications, Singapore, pp. 515–525 (December 2009)
20. Zhang, W., Che, D., Chen, Q., et al.: Development of gesture-changeable under-actuated humanoid robotic finger. *Chinese Journal of Mechanical Engineering* 23(2), 142–148 (2010)

# Empirical Copula Driven Hand Motion Recognition via Surface Electromyography Based Templates

Zhaojie Ju and Honghai Liu

Intelligent Systems & Biomedical Robotics Group, School of Creative Technologies,  
University of Portsmouth, Portsmouth, UK

**Abstract.** Current tendency of electromyography (EMG) based prosthetic hand is to enable the user to perform complex grasps or manipulations with natural muscle movements. In this paper, a new classifier is introduced to identify the naturally contracted surface EMG patterns for hand motion recognition. The recognition method utilizes a dependence structure as a motion template, which includes one-to-one correlations of surface EMG feature channels. Using an effective EMG feature, the proposed algorithm can successfully classify different complex motions from different subjects with a satisfactory recognition rate. To save the computational cost, re-sampling processing has been employed.

## 1 Introduction

An ideal active hand prosthesis, regarded as a sensible replacement of the lost hand, should be cheap, visually appealing, lightweight and long-running. More importantly it is supposed to be highly dexterous and easily controlled with sensory feedback. With the advances in the mechatronics, it has become feasible to build a prosthetic hand with emulational appearance, advanced materials and multiple active degrees of freedom. However, controlling the position and especially the exerted force of each finger can not be done naturally [1].

The surface electromyography (sEMG) is the current state-of-the-art technique for the control of active hand prostheses. The best known commercial hand prostheses are SensorHand and i-Limb. Both of them are limited to the degree of freedoms (DOF) or electrodes, which makes them far from the ideal active hand prosthesis [2]. Though others such as HIT/DLR prosthetic hand, CyberHand, APL-led Revolutionizing Prosthetic Arm and DEKA arm have more DOFs and electrodes, they are still suffering the problem of low sEMG recognition rates. How to enhance the sEMG recognition accuracy has become the main focus of the ongoing research of rehabilitation and prosthetics. The two key issues, feature selection and classifier design, play crucial roles in the performance of recognition algorithms. Feature selection is to define a feature vector from the original sEMG signal, while pattern classification is to discriminate feature vectors to different classes.

A variety of features have been extracted by different feature selection methods in the time and frequency domains. The time domain features of sEMG

include integral of sEMG (IEMG), wavelength (WL), variance (VIR). The frequency domain feature is a frequency domain short-term feature corresponding to the each frequency band. The zero crossing (ZC), slope sign changes (SSC), willison amplitude (WAMP), histogram of sEMG (HEMG), coefficients of autoregressive model (ARM) and wavelet features belong to frequency domain features, which are achieved by the time-frequency analysis methods such as short-time Fourier transform, wavelet transform and wavelet packet transform [2]; besides, Power Spectral Density (PSD) of sEMG signals can be estimated by Fourier transform methods, Yule-walker algorithm, Welch's method and the maximum entropy method. Liu *et al.* [3] has proven that ARM and HEMG are more effective than other six kinds of features including IEMG, WL, WAR, ZC, SSC and WAMP. However, in [4], eight features are tested with additive 50 Hz interference at various signal-to-noise ratios and WAMP outperforms other features including root mean square, WL, SSC, HEMG, median frequency, mean frequency, and ARM.

On the other hand, lots of classification methodologies have been proposed for processing and discriminating sEMG signals. As a computation technique that evolved from mathematical models of neurons and systems of neurons, Neural Network is becoming one of the most useful methods [5]. Tsenov *et al.* [6] proved that Multi-Layer Perception (MLP) gained a better recognition result than Radial Basis Function (RBF) and Learning Vector Quantization (LVQ). Naik *et al.* [7] improved the Back-Propagation Neural Network (BPN) using multi run ICA of sEMG to identify 6 hand gestures and achieved a relatively high classification accuracy. Other Neural Network based algorithms for classifying sEMG signals include Log-Linearized Gaussian Mixture network (LLGMN), Probabilistic Neural Network (PNN), Fuzzy Mean Max Neural Network (FMMNN) and Radial Basis Function Artificial Neural Network (RBFNN) [8]. The statistic classifiers such as Hidden Markov Model (HMM) [9], Gaussian Mixture Model (GMM) [10], Support Vector Machine (SVM) [11] and Beyes classifier [12] have also been used intensively in sEMG recognition; Additionally, fuzzy system has gained much attention in this research field [13]. A few studies has compared several different methods [11,12], e.g., Claudio *et al.* [1] reported SVM achieved a higher recognition rate than NN and Locally Weighted Projection Regression, while Liu [14] proposed Cascaded Kernel Learning Machine (CKLM) which has been compared to other classifiers such as k-nearest neighbor, multilayer NN and SVM. However, none of them has explained why the performance is enhanced and most of them are more concerned about individual muscle contraction than natural movement with combined multiple muscle contractions. In addition, there is a lack of consideration of sEMG's uncertainties such as non-stationary nature, muscle wasting, electrode position, different subjects and temperature impact. Muscle wasting or muscle fatigue can be considered as the decrease in the force generating capacity of a muscle and has been evidenced in numerous studies [15]. For the same hand motion, muscle fatigue results in a different sEMG signal which may cause a failure of the recognition method. Electrode position is crucial for the valid sEMG signal and leads to estimates of sEMG variables that

are different from those obtained in other nearby locations [16]. Temperature is also proved to have an important effect on the nerve conduction velocities and muscle actions [17]. These uncertainties need more consideration in extracting sEMG's features which are determinant to the performance of classifiers.

In this paper, we have applied a new method, Empirical Copula, to study the dependence structure of sEMG patterns with an effective sEMG feature, and then managed to build up the motion templates for recognizing complex hand motions. This paper is organized as follows: Section 2 describes recognition method using Empirical Copula, including the feature selection and re-sampling process. Section 3 gives the details of the experiment and demonstrates the recognition results. Finally the paper is concluded with concluding remarks.

## 2 Recognizing with Empirical Copula

Empirical Copula has successfully been used as a recognition algorithm for hand gestures captured from data glove in our previous study [18]. In this section, we will firstly revisit the theoretical foundation of Empirical Copula and dependence estimation, and then present the mechanism of the recognition algorithm.

### 2.1 Empirical Copula and Dependence Estimation

Empirical Copula is a characterization of the dependence function between variables based on observational data using order statistics theory and it can reproduce any pattern found in the observed data. If the marginal distributions are normalized, the Empirical Copula is the empirical distribution function for the joint distribution. Because only bivariate Empirical Copula will be employed in this paper (section 2.4), details of bivariate Empirical Copula is given as follows.

**Definition 2.11.** Let  $\{(x_k, y_k)\}_{k=1}^n$  denote a sample of size  $n$  from a continuous bivariate distribution. The Empirical Copula is the function  $C_n$  given by

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{\text{Num}((x, y) | x \leq x_{(i)}, y \leq y_{(j)})}{n} \quad (1)$$

where  $x_{(i)}$  and  $y_{(j)}$ ,  $1 \leq i, j \leq n$ , denote order statistics from the sample [19].

The Empirical Copula frequency  $c_n$  is given by

$$c_n\left(\frac{i}{n}, \frac{j}{n}\right) = \begin{cases} \frac{1}{n}, & \text{if } (x_{(i)}, y_{(j)}) \text{ is an element of the sample} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Note that  $C_n$  and  $c_n$  are related via

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \sum_{p=1}^i \sum_{q=1}^j c_n\left(\frac{p}{n}, \frac{q}{n}\right) \quad (3)$$

**Theorem 2.11.** Let  $C_n$  and  $c_n$  denote, respectively, the Empirical Copula and the Empirical Copula frequency function for the sample  $\{(x_k, y_k)_{k=1}^n\}$ . If  $\rho$  denotes the sample version of Spearman’s rho [20,21], then

$$\rho = \frac{12}{n^2-1} \sum_{i=1}^n \sum_{j=1}^n [C_n(\frac{i}{n} \cdot \frac{j}{n}) - \frac{i}{n} \cdot \frac{j}{n}] \tag{4}$$

Spearman’s rho is used to measure two variables’ association [19]. According to the definition and theorem, we can estimate one-to-one correlations between variables using Empirical Copula based on Spearman’s rho.

### 2.2 EMG Feature Extraction

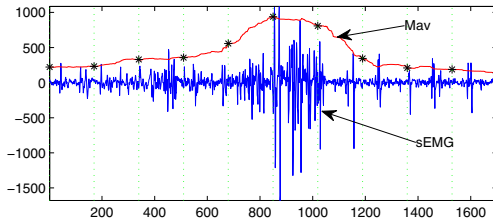
In this paper, one effective EMG feature, mean absolute value (MAV), is selected from EMG various features. An estimate of the mean absolute value of the signal,  $\bar{f}(x)$ , in segment  $x$  which is  $t_2 - t_1$  samples in length is given by

$$\bar{f}(x) = \frac{1}{t_2-t_1} \sum_{x=t_1}^{t_2} |f(x)| \tag{5}$$

where  $t_2 - t_1$  is the window size of the  $x$  segment. For example, figure 1 shows the MAV feature of the sEMG segment.

### 2.3 Re-sampling

Supposing the number of sampling points is  $n$  and number of attributes is  $m$ , for  $m \ll n$ , according to the equations 1 and 4, the time complexity of Spearman’s rho is  $O(n^3)$ . If number of samples is huge, computational time would be too long to get the result, which has been proven by our previous study [18]. To save the computational time, re-sampling processing is also employed in this study. We used the same strategy which is to take the samples at equal interval on the original sEMG features, e.g., in figure 1, sampling points of the MAV is set to be 11 instead of 1700, which alleviates the burden of the huge computational cost.



**Fig. 1.** Re-sample the data at equal interval

## 2.4 One-to-One Correlation and Motion Template

Supposing there are  $m$  variables which could be sEMG features (Equ. 5) of all channels in every motion,  $C_m^2$  is the total number of the one-to-one correlations. Let  $\rho_{ij}$  be the Spearman's rho between  $i$ th and  $j$ th variables, and the motion template is defined as the matrix  $\mathbf{P}$  of Spearman's rhos:

$$\mathbf{P} = \begin{pmatrix} \rho_{11} & \cdots & \rho_{1m} \\ \vdots & \ddots & \vdots \\ \rho_{m1} & \cdots & \rho_{mm} \end{pmatrix}$$

where  $\rho_{ij} = \rho_{ji}$  when  $i \neq j$  and  $\rho_{ij} = 1$  if  $i = j$ . Given  $s$  observations for one motion, the template is trained by taking the average of all Spearman's rho matrices.

$$\hat{\mathbf{P}} = \frac{\sum_{i=1}^s w_i \mathbf{P}_i}{\sum_{i=1}^s w_i} \quad (6)$$

where  $\mathbf{w} = [w_1, \dots, w_s]$  is a weight vector used to store the relative difference of each observation in the estimated template, so that more valid observation may carry larger weight than those with more uncertainties, which may be caused by noise, capturing devices, software and the environment. Figure 2 shows an example of the motion template representing the one-to-one correlations among the finger angles when grasping a book.

The matrix  $\mathbf{P}$  effectively aggregates the dependence relations of  $m$  variables into just one  $m \times m$  matrix, a highly reduced dimensionality of feature space. The relation matrix is naturally uniformed that the matrix is not dependent on differently sampled trials associated with specific speeds. This makes direct comparisons of relation matrices with differently sampled data feasible and computationally efficient.

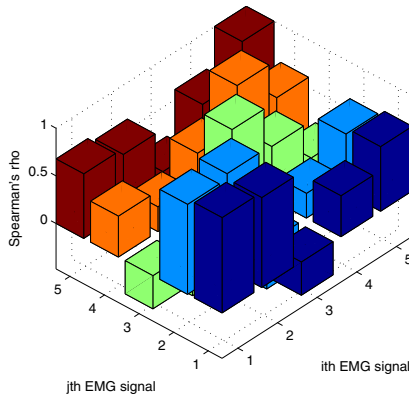


Fig. 2. An example of the sEMG motion template

## 2.5 Motion Recognition

Motion recognition is straightforward with the proposed template. It is achieved by finding the best match between an observed motion template and pre-trained motion templates. The proposed algorithm is applied on an observed motion to generate its motion template  $\mathbf{U} = \{\varrho_{ij}|i, j = 1, \dots, m\}$ . Its dissimilarity with the pre-trained template is achieved by

$$D_t = \|\mathbf{U} - \mathbf{P}\|_t = \left( \sum_{i=1}^m \sum_{j=1}^m |\rho_{ij} - \varrho_{ij}|^t \right)^{\frac{1}{t}} \quad (7)$$

$D_t$  is  $t$ -norm distance;  $t \geq 1$  and is a real number; usually we take  $t \in \{1, 2, \infty\}$  that  $D_1$  is the taxicab norm,  $D_2$  is the Euclidean norm and  $D_\infty$  is the maximum norm. The derived  $D_t$  norm infers the dissimilarity between the observed motion and the trained motions. The threshold of the template  $\mathbf{P}$  is defined as

$$th_{\mathbf{P}} = \frac{\sum_{i=1}^m \sum_{j=1}^m |\rho_{ij}|}{\alpha}; \quad (8)$$

where  $\alpha \geq 1$ , which indicates the threshold is  $100/\alpha$  percent of the whole absolute value of the template. Different datasets may have different  $\alpha$  values. In this paper, we set  $\alpha = 10$ , then the threshold of the template in fig 2 would be 8.33. The matching criterion of the motion recognition is that if  $D_t \leq th_{\mathbf{P}}$ , the observed motion is recognized as belonging to the trained motion.

## 3 Experiment

### 3.1 Subjects

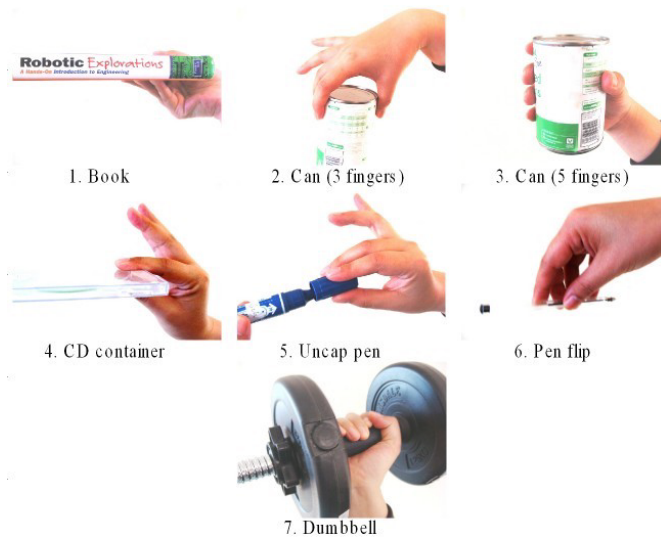
Eight (2 female, 6 male) healthy right-handed subjects volunteered for the study. Their ages range from 23 to 40 and average is 32.5 years; body height average is 175.5 cm; body mass average is 70 kg. All participants gave informed consent prior to the experiments and the ethical approval for the study was obtained from University of Portsmouth CCI Faculty Ethics Committee. All subjects were trained to manipulate different objects.

### 3.2 Tasks and Procedures

Participants had to perform 7 grasps or in-hand manipulations which are shown in figure 3 and the motions are listed as following

1. Grasp and lift a book using five fingers with the thumb abduction.
2. Grasp and lift a can full of rice using thumb, index finger and middle finger only.
3. Grasp and lift a can full of rice using five fingers with the thumb abduction.
4. Grasp and lift a disc container using thumb and index finger only.





**Fig. 3.** Selected hand tasks

5. Uncap and cap a marker pen using thumb, index finger and middle finger.
6. Pick up a pencil using five fingers, flip it and place it on the table.
7. Hold and lift a dumbbell.

The contact points had been decided for every object and every motion lasted about 2 to 4 seconds. Each motion was repeated 10 times. Between every two repetitions, participants had to relax the hand for 2 seconds in the intermediate state which is opening hand naturally without any muscle contraction. These intermediate states were used to segment the motions. Once one motion with ten repetitions was finished, participants had to relax the hand for 2 minutes before the next motion started. This was designed to overcome the effects of muscle fatigue.

### 3.3 Data Collection

The sEMG of 5 forearm muscles, i.e. flexor carpi radialis, flexor carpi ulnaris, flexor pollicis longus, flexor digitorum profundus and extensor digitorum, were measured using DataLINK system from Biometrics LTD with a gel-skin contact area of about 4 cm<sup>2</sup> for each bipolar electrode and a centre to centre recording distance of 20 mm. The sampling frequency of DataLINK system in our experiment was set to be 1000Hz and sEMG signals were amplified 1000 times and bandwidth is 20 to 460 Hz using a sEMG amplifier (SX230FW sEMG Amplifier, Biometrics LTD). To obtain clearer signals, subjects were scrubbed with alcohol and shaved if necessary and then electrodes were applied over the body using the die cut medical grade double sided adhesive tape. Electrodes locations were selected according to the Musculoskelet of these five muscles and confirmed

by muscle specific contractions, which include manually resisted finger flexion, extension and abduction. The real time sEMG signals were visualized on a computer screen giving participants feedback to choose the positions of electrodes with stronger sEMG signals.

### 3.4 Results and Discussion

All original sEMG signals were separated by deleting the intermediate states. According to equation 5, mean absolute value was calculated with window size 400 for every motion and re-sampled to 100 sampling points. Half of the mean absolute value were used to train the Empirical Copula and the others were used for testing, according to the section 2. Table 1 presents the overall results in terms of different motions and different subjects. ‘S1’ means the subject 1; ‘M1’ denotes motion one; ‘SR’ is the recognition rate of one particular subject and ‘MR’ is the recognition rate for one particular motion. The number from 1 to 5 represents the number of the correctly identified motions.

**Table 1.** Recognition rates in terms of different motions and different subjects

	M1	M2	M3	M4	M5	M6	M7	SR
S1	5	2	4	2	5	3	5	74.29%
S2	5	5	5	4	5	4	5	94.29%
S3	3	4	5	5	4	3	4	80%
S4	2	5	5	4	5	5	5	88.57%
S5	3	5	4	1	4	4	5	74.29%
S6	2	5	5	4	4	4	5	82.86%
S7	5	4	4	4	5	5	5	91.43%
S8	4	5	2	5	5	5	5	88.57%
MR	72.5%	87.5%	85%	72.5%	92.5%	82.5%	97.5%	<b>84.29%</b>

From the table, regarding different motions from all subjects, the M7 gets the highest recognition rate, 97.5%, with only one motion wrongly recognized from S3, while the M5 achieves the second highest, 92.5%. These two motions need much more force requiring stronger muscle contraction than others, which makes the signals more identifiable than others. M1 and M4 have the lowest rate, 72.5%, and about half mis-identified motions of them are caused by M3 and M2 respectively. This is mainly because of the similarity between M1 and M3 and the similarity between M2 and M4, which is confirmed from the figure 3. Needing to point out that the M4 from S5 scores only 1 out of 5 which is the lowest in all the motions.

On the other hand, for different subjects doing all motions, the recognition rates range from 74.29% to 94.29%. For S2, only 2 motions have been wrongly identified and they are from M4 and M5. Except S1 and S5, whose rates are both 74.29%, others are all equal or above 80%. In the whole 56 motions (7 motions multiple 8 subjects), more than half score 5 out of 5 correctly identified.

In addition, the overall recognition rate reaches 84.29%. These have proved that the recognition algorithm is an effective classifier of sEMG signals to identify different hand motions from different operators.

## 4 Concluding Remarks

sEMG signal recognition has been intensively studied aiming to find a reliable classifier for the natural complex motions. In this paper, Empirical Copula has been applied to study the dependence structure of sEMG patterns with an effective feature. The proposed motion template composed of one-to-one correlations of the sEMG feature channels has been used to classify the natural complex hand motions. This template frees the limitation of the motion speed when comparing different motions, so that recognition process is easily achieved by finding the dissimilarity between an observed motion template and pre-trained motion template. The experiments on grasp tasks or in-hand manipulations from different subjects showed that the classifier of Empirical Copula can identify complex motions with an overall 84.29% recognition rate. In addition, huge computational cost of the algorithm has been saved by the proposed re-sampling pre-processing. In the future, the recognition performance of this approach can be improved by using another sEMG feature or combined features, such as WL, VIR, WAMP and so on. It would be more convincing that it could be compared with other dominant classifiers such as neural networks and HMM within the same condition. Another future research direction is to study sEMG patterns constructed by fusing qualitative description for complex unconstrained motions [22,23].

## References

1. Castellini, C., van der Smagt, P.: Surface EMG in advanced hand prosthetics. *Biological Cybernetics* 100(1), 35–47 (2009)
2. Chu, J.U., Moon, I., Lee, Y.J., Kim, S.K., Mun, M.S.: A supervised feature-projection-based real-time EMG pattern recognition for multifunction myoelectric hand control. *IEEE/ASME Transactions on Mechatronics* 12(3), 282–290 (2007)
3. Huang, H.P., Liu, Y.H., Wong, C.S.: Automatic EMG feature evaluation for controlling a prosthetic hand Using a supervised Feature Mining Method: An Intelligent Approach. In: *IEEE International Conference on Robotics and Automation*, vol. 1, pp. 220–225. IEEE, Los Alamitos (1999/2003)
4. Phinyomark, A., Limsakul, C., Phukpattaranont, P.: Emg feature extraction for tolerance of 50 hz interference. In: *Proc. of PSU-UNS Inter. Conf. on Engineering Technologies, ICET 2009*, pp. 289–293 (2009)
5. Ahsan, M.R., Ibrahimy, M.I., Khalifa, O.O.: EMG Signal Classification for Human Computer Interaction: A Review. *European Journal of Scientific Research* 33(3), 480–501 (2009)
6. Tsenov, G., Zeghibib, A.H., Palis, F., Shoylev, N., Mladenov, V.: Neural networks for online classification of hand and finger movements using surface EMG signals. In: *8th Seminar on Neural Network Applications in Electrical Engineering*, pp. 167–171 (2006)

7. Naik, G.R., Kumar, D.K., Palaniswami, M.: Multi run ICA and surface EMG based signal processing system for recognising hand gestures. In: 8th IEEE International Conference on Computer and Information Technology, CIT 2008, pp. 700–705 (2008)
8. Mobasser, F., Hashtrudi-Zaad, K.: A method for online estimation of human arm dynamics. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 1, p. 2412 (2006)
9. Chan, A.D.C., Englehart, K.B.: Continuous myoelectric control for powered prostheses using hidden Markov models. *IEEE Transactions on Biomedical Engineering* 52(1), 121–124 (2005)
10. Huang, Y., Englehart, K.B., Hudgins, B., Chan, A.D.C.: Optimized Gaussian mixture models for upper limb motion classification. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 1 (2004)
11. Kawano, S., Okumura, D., Tamura, H., Tanaka, H., Tanno, K.: Online learning method using support vector machine for surface-electromyogram recognition. *Artificial Life and Robotics* 13(2), 483–487 (2009)
12. Englehart, K., Hudgins, B.: A robust, real-time control scheme for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering* 50(7), 848–854 (2003)
13. Khezri, M., Jahed, M.: Real-time intelligent pattern recognition algorithm for surface EMG signals. *BioMedical Engineering OnLine* 6(1), 45 (2007)
14. Liu, Y.H., Huang, H.P., Weng, C.H.: Recognition of electromyographic signals using cascaded kernel learning machine. *IEEE/ASME Transactions on Mechatronics* 12(3), 253–264 (2007)
15. Wang, G., Ren, X., Li, L., Wang, Z.: Multifractal analysis of surface EMG signals for assessing muscle fatigue during static contractions. *Journal of Zhejiang University-Science A* 8(6), 910–915 (2007)
16. Mesin, L., Merletti, R., Rainoldi, A.: Surface EMG: The issue of electrode location. *Journal of Electromyography and Kinesiology* 19(5), 719–726 (2009)
17. Feinberg, J.: EMG: Myths and Facts. *HSS Journal* 2(1), 19–21 (2006)
18. Ju, Z., Liu, H.: Recognizing Hand Grasp and Manipulation through Empirical Copula. *International Journal of Social Robotics* (2010) (awaiting publishing)
19. Nelsen, R.B.: An introduction to copulas. Springer, Heidelberg (2006)
20. Kruskal, W.H.: Ordinal measures of association. *Journal of the American Statistical Association* 53(284), 814–861 (1958)
21. Lehmann, E.L.: Some concepts of dependence. *The Annals of Mathematical Statistics* 37(5), 1137–1153 (1966)
22. Liu, H., Brown, D.J., Coghill, G.M.: Fuzzy qualitative robot kinematics. *IEEE Transactions on Fuzzy Systems* 16(3), 802–822 (2008)
23. Liu, H.: A fuzzy qualitative framework for connecting robot qualitative and quantitative representations. *IEEE Transactions on Fuzzy Systems* 16(6), 1522–1530 (2008)

# A Continuous Control Scheme for Multifunctional Robotic Arm with Surface EMG Signal

Pinghua Hu, Shunchong Li, Xinpu Chen, Dingguo Zhang, and Xiangyang Zhu

State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering,  
Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China  
hph1923@sjtu.edu.cn

**Abstract.** This paper applies real time pattern recognition into the control of robotic hand with surface electromyographic (sEMG) signal. We focus on the hardware system design and the control strategy implementation. Time domain statistic methods are employed to extract the features, which have good effects on the pattern recognition. After the feature dimension reduction by Fisher linear discriminant (FLD), the feature vector is classified by a multi-layer perception (MLP) neural network. At last the data of several subjects is analyzed, and it shows good recognition accuracy. Using this system, the subjects can control a robotic arm to perform desired movements intuitively.

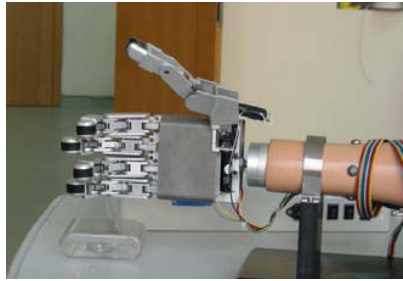
**Keywords:** surface electromyographic (sEMG) signal, classification, robotic arm, continuous control, DSP.

## 1 Introduction

Electromyographic (EMG) signal is a kind of complex signal which is generated by the contraction of muscles. EMG signal contains much information of the muscle activity and it can be collected on the skin surface using electrodes. As a good control resource, it has been widely used in many applications. For example, EMG signal can be used to control prosthesis [1-3], robot arm [5], computer input device [11], wheelchair [12] and so on. In these applications, EMG signal processing is a key factor. Feature extraction and feature classification are two important procedures, and both of them may affect the final performance significantly. The effort to develop advanced pattern recognition algorithm was largely limited by the computing capacity of the day when pattern recognition was firstly applied to EMG control. Accompanied by the ever-growing computing capacity of CPU, more and more advanced pattern recognition algorithms have been developed. For feature extraction, there are time domain statistics [1], autoregressive coefficient [2], cepstral coefficient [3], and so on. For pattern classification, there are artificial neural networks [4], genetic algorithms [5], fuzzy logic classifiers [6], [7], linear discriminant analysis [8] and so on.

Among the various applications of EMG, we are interested in control of robot arm and prosthetic hand as they share some common grounds. Although EMG controlled prosthetic hand emerged in 1970s [2], there is still one big challenge to develop a more intuitive and functional prosthetic hand at present, i.e. the available signal processing for sEMG only allows sequential and on/off control for a prosthetic hand with

more than one degree of freedom (DOF). The mature products in the market are still lacked of multifunction, for example, the famous prosthetic hand of Otto Bock has only one DOF. In this work, we aim to apply a continuous real-time control scheme on a robotic arm with multi-DOFs, and illustrate the feasibility of future application in prosthesis. Two classic methods, time domain statistics (TDS) and multiple layers perception (MLP), were employed in preliminary experiments. Actually, we have already designed a kind of prosthetic hand named as SJTU hand II (see Fig.1), which is based on the on/off control. The new version (SJTU hand III) is improved to adopt continuous control scheme, but it is still under manufacturing. In this work, we test our continuous control scheme on a robotic arm instead, which has the same control protocol in the hard-ware level as that for the prosthetic hand.



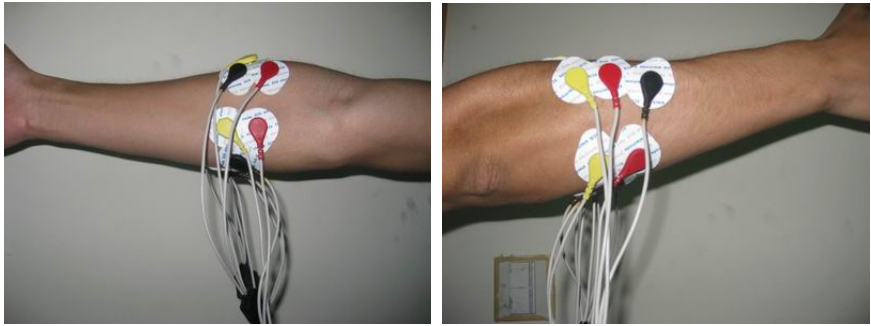
**Fig. 1.** Actual picture of the prosthetic hand (SJTU hand II)

## 2 Methodology

### 2.1 Data Acquisition

We collect the experimental data using our self-designed EMG system. In this work, surface EMG signals are acquired through four channels on the concerned four muscle groups (palmaris longus, flexor carpi ulnaris, flexor digitorum superficialis, extensor digitorum) of forearm as shown in Fig.2. Each channel consists of three electrodes: two of them for data acquisition and the residual one for reference. The subjects are asked to practice four contractions (hand open, hand close, flexion, and extension) before acquiring the data. The skin areas of interest are abraded with alcohol before attaching the electrodes.

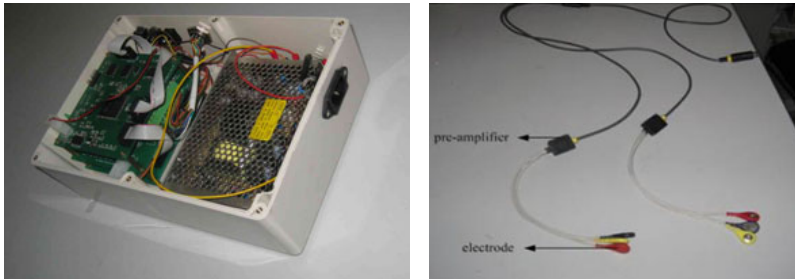
The EMG signal is processed by a two-step amplifier, with a pre-amplifier of 100 times and an adjustable post-amplifier of 2~10 times. The first amplification is set in the middle of the cable, and the second amplifier is set in the control box (see Fig.3). After pre-amplification there is a long distance to transmit the signal to the control box, so the signal must be strong enough to pass through the cable. For this reason we finally set the pre-amplification time at 100 according to many trial-and-error experiments. It's critical to shield the signal at the first amplification for the large number of amplification. After filtering, the EMG signal is collected by DSP with a 12 bit A/D



(a)

(b)

**Fig. 2.** Placement of electrodes (a) posterior view (palmaris longus, flexor carpiulnaris) (b) anterior view (flexor digitorum supercifialis, extensor digitorum)



(a)

(b)

**Fig. 3.** Hardware system. (a) Electronic control box(b) Cable with electrodes.

converter under the frequency of 1000HZ, as the useful sEMG signal ranges from the frequency of 10—500HZ. A feature set was extracted from the sEMG signal every 300 ms (namely 300 A/D samples). Breakpoints are set in the DSP so that the PC can read the data sets every 200 time windows from the RAM of DSP, and the subjects can relax his muscle. The data is processed by MATLAB. Finally, the coefficient matrices are provided to the DSP for pattern classification. The whole data acquisition process takes only 15 minutes. The whole control process is illustrated in Fig. 4.

For DSP component, we adopt TMS320F2812 (Texas Instruments Incorporated), which has 8 independent PWM outputs. Since it only needs to write new data in relative registers to change the width of the pulse of PWM, it consumes little resource of the CPU to control the motor. The EMG processing algorithms are implemented in this DSP chip, which helps a lot to control a responsive robot arm or prosthetic hand. In the electronic control box as shown in Fig.3 (a), there are some other auxiliary circuits such as oscillation circuit, ADC input interface, PWM output interface, JTAG regulation interface and expanded external RAM.

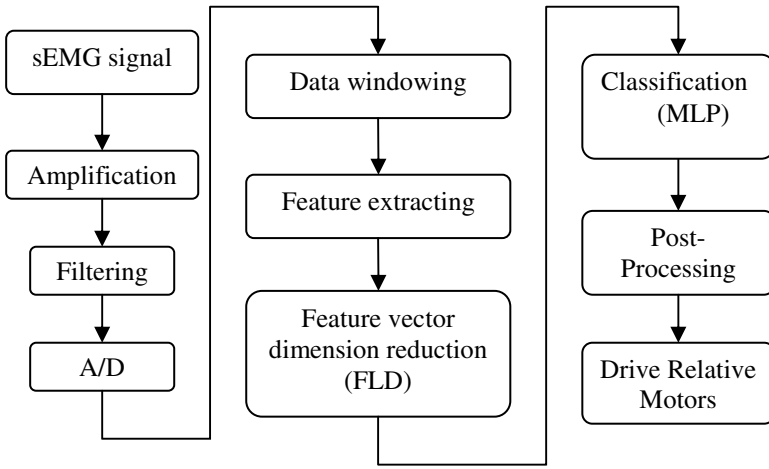


Fig. 4. Data flow of control process

## 2.2 Data Processing

EMG data processing is the kernel of the control system. The flowchart of EMG process is shown in Fig.5. It consists of four stages: analysis window, TDS, FLD, and MLP. In the following, we discuss them respectively.

### Analysis Window Selection

The length of the analysis window and the window increments are important for continuous EMG control [8]. The two factors are critical to make a trade-off between classification accuracy and the response time. After many experiments we set the window length to 300ms and the window increment to 100ms. The feature set consists of four time domain statistics for each channel (the number of zero crossings, the waveform length, the number of changes in slope sign, and the mean absolute value) which has shown satisfactory performance in the transient signal classification [1]. In a continuous classification scheme, the data are essentially stationary in every single data window, so it's feasible to use these time domain statistics here.

### Time Domain Statistics (TDS)

**Mean Absolute Value:** An estimate of the mean absolute value of the signal in segments, which are samples in length, is given by

$$\bar{x}_i = \frac{1}{L} \sum_{k=1}^L |x_k| \quad \text{for } i=1, \dots, I. \quad (1)$$

where  $x_k$  is the  $k$ -th sample in window  $i$ , and  $I$  is the total number of windows in the records.



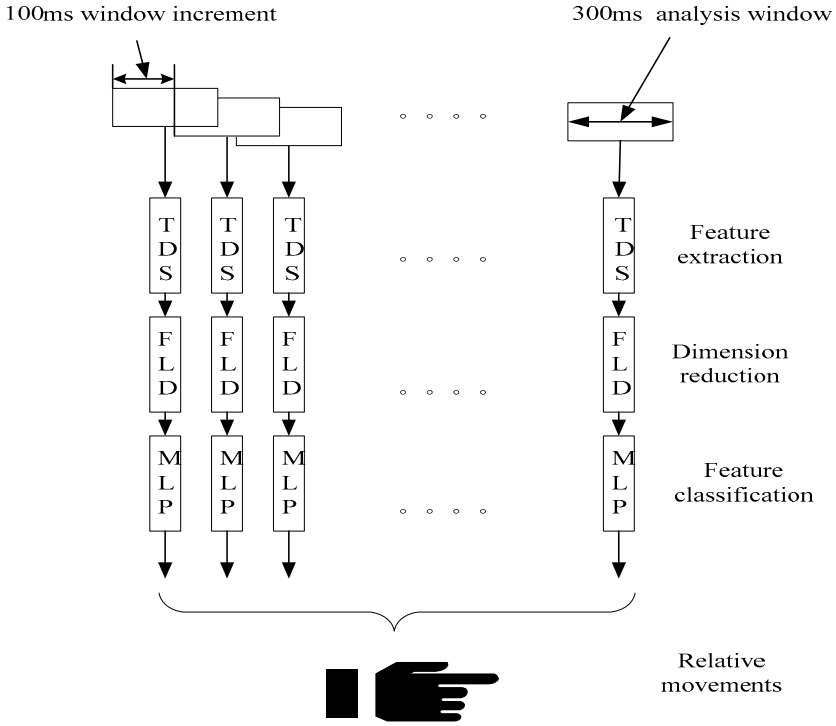


Fig. 5. Flow chart of EMG data processing

**Zero Crossings:** A simple frequency measure can be obtained by counting the number of times the waveform crosses zero. A threshold ( $\epsilon$ ) must be included in the zero crossing calculation in order to reduce the zero crossings induced by noise. Given two consecutive samples  $x_k$  and  $x_{k+1}$ , the zero crossing count is increased, if

$$\{x_k > 0 \& x_{k+1} < 0\} \text{ or } \{x_k < 0 \& x_{k+1} > 0\} \quad (2)$$

and

$$|x_k - x_{k+1}| \geq \epsilon \quad (3)$$

**Changes in Slope Sign:** A feature that may provide another measure of frequency content is the number of changes in the slope sign. In addition, a suitable threshold must be chosen to reduce the changes of slope sign induced by noise. Given three consecutive samples,  $x_{k-1}$ ,  $x_k$  and  $x_{k+1}$ , the slope sign change is increased if

$$\{x_k > x_{k-1} \& x_k > x_{k+1}\} \text{ or } \{x_k < x_{k-1} \& x_k < x_{k+1}\} \quad (4)$$

and

$$|x_k - x_{k+1}| \geq \epsilon \text{ or } |x_k - x_{k-1}| \geq \epsilon. \quad (5)$$

**Waveform Length:** A feature that provides information on the waveform complexity in each segment is the waveform length. Simply it is the cumulative length of the waveform over the segment, and defined as:

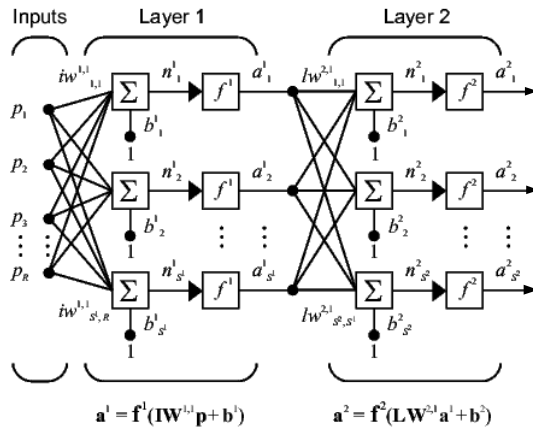
$$l_0 = \sum_{k=1}^L |\Delta x_k|. \quad (6)$$

where  $\Delta x_k = x_k - x_{k-1}$ . The resultant values indicate a measurement of waveform amplitude and frequency. The durations are all within a single parameter. The feature sets are then concatenated to form 300 16-dimensional feature vectors for each contraction. Before using a multi-layer perceptron (MLP) network, the dimension of the feature vector is reduced to 3 by Fisher linear discriminant (FLD). Then the final decision is made to drive the robotic arm. FLD is represented by

$$\eta_i = WX_i \quad (7)$$

where  $X_i$  is the 16-dimensional feature vector extracted from an analysis window,  $W$  is the linear projection matrix.  $\eta_i$  is the vector after dimension reduction. More details of FLD can be found in [9].

**MLP Neural Network:** Here, a two-layer neural network is adopted as a classifier, and the structure is shown in Fig.6. In this network, each element of the input vector  $p$  is connected to each neuron input through the weight matrix  $W^l$  of first layer. The  $i$ -th neuron has a summer that gathers its weighted inputs and bias to form its own scalar output  $n_i^1$ . The variables  $n_i^1$  are taken together to form an  $S$ -element net input



**Fig. 6.** A two-layer MLP network with  $R$  input elements and  $S^l$  neurons

vector  $n$ . Finally, the outputs of neuron layer 1 form a column vector  $a^1$ . Then  $a^1$  acts as the input vector of layer 2 (output layer).  $a^2$  is the output vector of the network.

$$a^1 = f^1(W^1 p + b^1) . \quad (8)$$

$$a^2 = f^2(W^2 a^1 + b^2) . \quad (9)$$

$$f^1(x) = \frac{1}{1 + e^{-x}} . \quad (10)$$

where  $f^2$  is a hard-limit transfer function, which returns 0 or 1.

### 2.3 Robotic Arm

A general robotic arm in lab teaching is used as a platform to test the proposed EMG control system. The robotic arm has five DOFs, and five motors are fixed on corresponding joints (see Fig.7). No.1 motor drives the arm to rotate in the horizontal plane, No.2 motor makes it move in the vertical plane. No.3 and No.4 motors drive the arm to move around their own axes respectively. No.5 motor works to control the hand with two movements: hand open and hand close. In this work, we only need two DOFs, so we fixed joint 1, 3, and 4. We use joint 2 to mimic the wrist motions: flexion and extension. The motors are driven by two channels of PWM outputs. Under the source voltage of 5V, the motor can rotate at a speed of 200ms/60°.

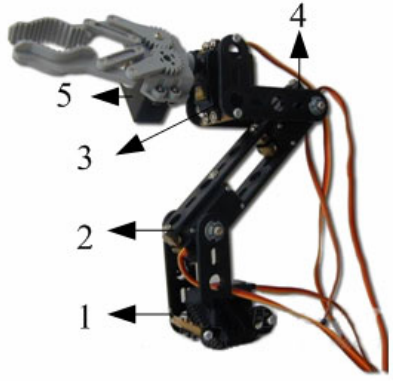


Fig. 7. Robotic arm with five DOFs. The digital number indicates the position of motors.

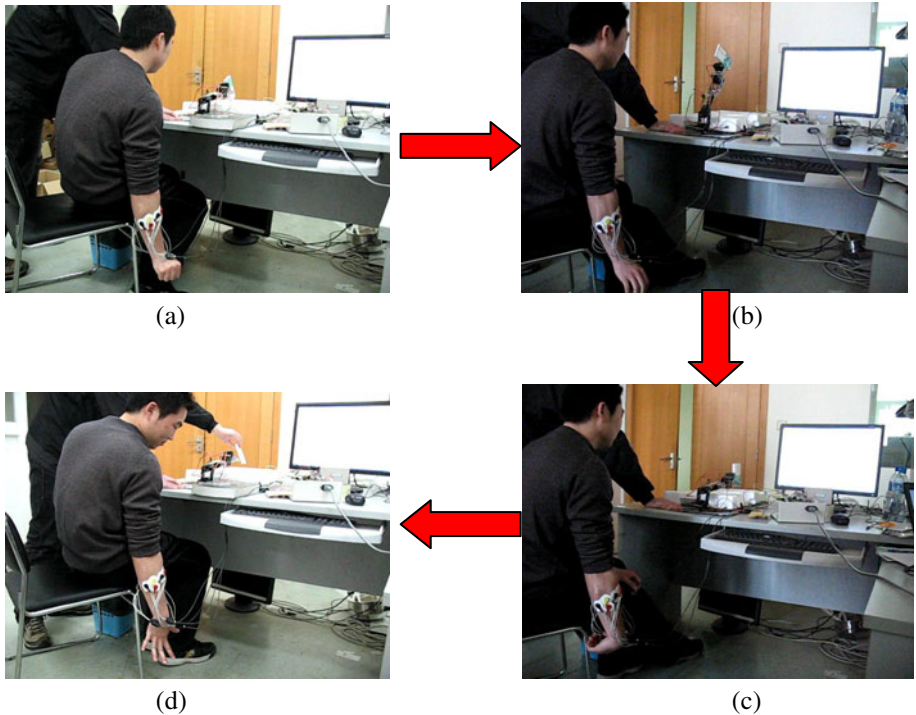
## 3 Result

Table 1 shows the results of 5 subjects' off-line classification accuracy in this control scheme, which can work as a reference to the on-line mode accuracy. The discriminant difference between off-line mode and on-line mode was caused by the different CPU performance of the PC and DSP. But the 32-bit CPU of the DSP is enough to

**Table 1.** Classification accuracy of four motions

	Subject.1 (%)	Subject.2 (%)	Subject.3 (%)	Subject.4 (%)	Subject.5 (%)
Hand close	91.5	94	93	97.5	96.5
Hand open	97	94	92.5	95.5	94.5
Flexion	94	100	99	99	99
Extention	98	99.5	99	99	97.25
Mean	94.625	96.875	95.875	97.25	97.25

gain a good performance on addition and multiplication, which are the two operations in this work. We collect 500 feature vectors during each of the four contractions, and 200 vectors were use for MLP training. The rest 300 vectors are used to test the classification accuracy. The whole calculation process is implemented on MATLAB.



**Fig. 8.** Experimental demonstration of robotic arm performing continuous movements under the subject’s EMG control. The red big arrows indicate the sequence of movements. (a) grasping object, (b) lifting arm, (c) lowering arm, (d) opening hand.

We test the performance of our control scheme on the robotic arm as follows. A healthy subject uses his EMG signals to control the robotic arm. In the demonstration, the arm can do four continuous motions according to the subject's intention: seize an object, then move it to a higher place, lower the arm, and drop it at last (see Fig.8). The accuracy of this movement set is about 75%. The EMG signals acquired via our self-designed EMG system are easily to be affected by the environment noise, which is the fatal factor affecting the accuracy. So the result is not very ideal. If EMG signals are acquired using the commercial product, i.e. Mega system (ME6000), the classification accuracy is higher than that in Table 1. But the problem is that Mega system is not compatible with our system, and it cannot be redesigned.

## 4 Discussion

At the beginning of the contraction switching, there are always some noises (pulse signals) that are much stronger than the useful signal. We will introduce a majority vote function to prohibit the influence brought by the noises. During one decision, six modes are put into the function, and the mode with the biggest number is the final decision. If there are the same numbers of modes, the final decision is the previous sample.

Although TDS performs very well in the feature extraction, a better recognition accuracy may be gained by using discriminant Fourier-derived cepstrum (DFC) [10]. We will compare these two methods in the future work.

Actually, more than 4 kinds of contractions can be classified by this scheme, but the accuracy deteriorates much unless more EMG channels are used. Therefore, perhaps DFC may solve this problem, otherwise, more channels must be provided to ensure the accuracy at a high level.

In near future, we will realize the continuous control of the prosthetic hand (SJTU hand III). Five DC motors are employed. The index finger and middle finger are driven by one motor, and the ring finger and little finger are driven by one motor. These two motors can realize the flexion and extension of relative fingers. The thumb is driven by two motors: one of them to realize the flexion and extension, the other to realize the retroflexion. The wrist joint is driven by one motor to realize supination and pronation. An underactuated mechanism is applied to design the prosthetic hand that is less motors are employed to drive a larger number of DOFs. Using this structure, the hand can grasp objects naturally.

A visual feedback can be applied so that the disabled can use the prosthetic hand more intuitively. It acts in such a way: a movement produced by the prosthetic hand can be programmed to finish in several steps, and the amputee can make it stop at any step when grasping different objects. Some force sensors can be fixed on the fingers to make the control system more dimensional. To accomplish this task, much investigation was needed to make the hand acting smoothly.

Some clinical study is undergoing now. We need to do more experiments on the amputees to learn the muscle characters of the deficit arm, and then conceive an effective way for the amputees to use the phantom hand.

Although our current project intends to implement the proposed EMG control strategy in prosthesis, obviously the method is in nature very suitable for teleoperation of

a robot arm in telemedicine, aerospace, or some hazardous environments and so on. So the scope of the possible applications is quite wide indeed.

## 5 Conclusion

This paper provides a robust, multifunctional and responsive way to control the robotic arm using surface EMG signals. EMG signals are acquired from four channels on the forearm muscles. The EMG processing includes TDS for feature extraction, FLD for dimension reduction, and MLP for classification. In this paradigm, robotic arm can realize four continuous movements under the subject's EMG control. The result is acceptable. This control system was designed in a distributed structure so that the hand can be controlled by other signals such as electroencephalogram (EEG) if needed. This work lays a foundation for our prosthetic hand control in future.

## Acknowledgment

This work was supported by the National High Technology Research and Development Program of China (No. 2009AA04Z212), and Research Fund for the Doctoral Program of Higher Education of China (No. 20090073120047).

## References

1. Hudgins, B., Parker, P.A., Scott, R.: A new strategy for multifunction myoelectric control. *IEEE Trans. Biomed. Eng.* 40(1), 82–94 (1993)
2. Graupe, D., Cline, W.: Functional separation of EMG signals via ARMA identification methods for prosthesis control purposes. *IEEE Trans. Syst. Man Cybern.* 5, 252–259 (1975)
3. Kang, W., Shiu, J., Cheng, C., Lai, J., Tsao, H., Kuo, T.: The application of cepstral coefficients and maximum likelihood method in EMG pattern recognition (movements classification). *IEEE Trans. Biomed. Eng.* 42, 777–785 (1995)
4. Englehart, K., Hudgins, B., Stevenson, M., Parker, P.A.: Classification of transient myoelectric signals using a dynamic feedforward neural network. In: *Proc. World Cong. Neural Networks*, Washington, DC (July 1995)
5. Farry, K., Walker, I.D., Baraniuk, R.G.: Myoelectric teleoperation of a complex robotic hand. *IEEE Trans. Robot. Autom.* 12(5), 775–788 (1996)
6. Leowinata, S., Hudgins, B., Parker, P.A.: A multifunction myoelectric control strategy using an array of electrodes. In: *The 16th Annual Congress of the International Society Electrophysiology and Kinesiology*, Montreal, PQ, Canada (1998)
7. Weir, R.F.f., Ajiboye, A.B.: A multifunction prosthesis controller based on fuzzy-logic techniques. Presented at the 25th Silver Anniversary International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), Cancun, Mexico, September 17-21 (2003)
8. Englehart, K., Hudgins, B.: A robust, real-time control scheme for multifunction myoelectric control. *IEEE Trans. Biomed. Eng.* 50(7), 848–854 (2003)
9. Duda, R., Hart, P., Stork, D.: *Pattern Classification and Scene Analysis*. Wiley, New York (1973)

10. Chen, X.P., Zhu, X.Y., Zhang, D.G.: Use of the discriminant fourier-derived cepstrum with feature-level post-processing for the surface electromyographic signal classification. *Physiological Measurement* 30, 1399–1413 (2009)
11. Wheeler, K.R., Jorgensen, C.C.: Gestures as input: Neuroelectric joysticks and keyboards. *IEEE Pervasive Computing* 2(2), 56–61 (2003)
12. Choi, K., Sato, M., Koike, Y.: A new, human-centered wheelchair system controlled by the EMG signal. In: *International Joint Conference on Neural Networks (IJCNN)*, Vancouver, pp. 4664–4671 (2006)

# On the Robust Control of Systems Preceded by Differential Equation-Based Hysteresis Nonlinearities

Ying Feng<sup>1,2</sup>, Juan Du<sup>1</sup>, and Chun-Yi Su<sup>2</sup>

<sup>1</sup> College of Automation Science and Engineering,  
South China University of Technology, Guangzhou, Guangdong, 510641, China

<sup>2</sup> Department of Mechanical and Industrial Engineering,  
Concordia University, Montreal, Quebec, H3G 1M8, Canada

**Abstract.** In this paper, a robust control approach for a class of nonlinear systems preceded by unknown hysteresis nonlinearities is addressed. Due to the complexity of the hysteresis characteristics, the hysteresis can not be linearized directly, and the effects caused by the hysteresis will degrade the system performance. Therefore, it is necessary to design an effective controller mitigating the negative effects. In this paper, the unknown hysteresis is represented by a differential equation-based hysteresis model - Duhem model. By exploring the characteristics of the Duhem model, the developed robust controller ensures the global stability of the system without constructing the hysteresis inverse. The effectiveness of the proposed control approach is demonstrated through a simulation example.

## 1 Introduction

Hysteresis is a nonlinear phenomenon that appears in various systems including ferromagnetic materials, mechanical actuators and electronic relay circuits. For example, the smart material based actuators utilized in the high precision positioning inevitably exhibit inherent hysteretic behaviors, which limits the position precision. Generally speaking, when the systems are preceded by unknown hysteresis nonlinearities, the system performance will be degraded, causing the undesirable inaccuracy or oscillations of the systems. Consequently, development of a control technique to overcome the limitations introduced by the hysteresis nonlinearity for the control system has currently attracted significant interest [3, 17-25].

As one of non-smooth nonlinearities, hysteresis has special characteristics, such as the nondifferentiation and multivaluedness etc., which makes it difficult to linearize the hysteresis nonlinearities directly. Therefore, the foremost step for the controller design of the preceded hysteretic systems is to characterize these hysteresis nonlinearities thoroughly.

Addressing this task, numerous hysteresis models have been developed to express various hysteresis phenomena. They can be roughly classified as operator-based hysteresis models [2, 8, 13, 23] and differential equation-based hysteresis



models [4] [5] [6] [14] [15]. Based on these hysteresis modelling approaches, some control approaches have been developed in the literature to mitigate the effects caused by the uncertain hysteresis input. The obstacle for the controller design of such hysteretic systems is the immeasurability of the hysteresis output (system input), which causes available control schemes not to be applied directly.

Focusing on such a cascade hysteretic system structure, the control frames can be also roughly classified as constructing the hysteresis inverse [3] [9] [11] [20] and without constructing the hysteresis inverse [18] [25]. By constructing the hysteresis inverse as compensator, the available control methods can be utilized directly. However, it may cause the complexity of the system structure and the difficulty for the system closed-loop stability analysis. Alternatively, another control scheme without the hysteresis inverse tries to avoid the complexity of the closed-loop system stability analysis. Since this scheme does not need the hysteresis inverse, the developed control methods should be combined with the characteristics of the hysteresis models, which is the main feature of this control scheme.

In this paper, a differential equation-based hysteresis model, Duhem model [5] [6] [12], is adopted to describe the unknown hysteresis. The Duhem model can describe a class of general hysteresis shapes by choosing different input functions. However, due to the existence of the nonlinear input functions, it generates the difficulty for the controller design, which needs a special new treatment. By exploring the characteristics of the Duhem model, a robust control approach for a class of nonlinear systems with hysteresis input is proposed in this paper. Without constructing the hysteresis inverse, the proposed robust control algorithm ensures the global boundedness of the closed-loop system and achieves the desired tracking precision. Simulation results are provided to show the effectiveness of the proposed approach.

This paper is organized as follows: Section 2 states the control problem. Section 3 introduces the differential equation-based hysteresis model - Duhem model briefly. Section 4 presents the robust controller and analyzes the stability and performance of the closed-loop system. To demonstrate the effectiveness of the proposed approach, simulation results are presented in Section 5. Finally, Section 6 concludes this paper.

## 2 Problem Statement

In this paper, a class of SISO nonlinear systems in [16] [18] [19] preceded by unknown hysteresis are considered.

$$x^{(n)}(t) + \sum_{i=1}^k a_i Y_i(x(t), \dot{x}(t), \dots, x^{(n-1)}(t)) = bw(t) \quad (1)$$

where  $Y_i$  are known continuous linear or nonlinear functions, parameters  $a_i$  are unknown constants and control gain  $b$  is unknown bounded constant. It is a common assumption that the sign of  $b$  is known. Without losing generality, we

assume  $b > 0$ .  $w$  is the input of the system. Also,  $w$  is the output of the hysteresis, which can be defined as

$$w(t) = P[v](t) \quad (2)$$

with  $v(t)$  as the input of hysteresis.

The control objective is to design a control law for  $v(t)$ , forcing the plant state  $\mathbf{x}(t)$  to follow a specified desired trajectory,  $\mathbf{x}_d(t)$ , i.e.,  $\mathbf{x}(t) \rightarrow \mathbf{x}_d(t)$  as  $t \rightarrow \infty$ . The desired trajectory is subjected to the following assumption.

*Assumption 1.* The desired trajectory  $\mathbf{x}_d = [x_d, \dot{x}_d, \dots, x_d^{(n-1)}]^T$  is continuous. Furthermore,  $[x_d^T, \dot{x}_d^{(n)}]^T \in \Omega_d \subset R^{n+1}$  with  $\Omega_d$  being a compact set.

### 3 Duhem Hysteresis Model

In this paper, the unknown hysteresis is represented by Duhem model, which belongs to differential equation-based hysteresis models. In this section, the properties of Duhem model is introduced briefly.

For the Duhem model, it can be defined as

$$\frac{dw}{dt} = \alpha \left| \frac{dv}{dt} \right| [f(v) - w] + \frac{dv}{dt} g(v) \quad (3)$$

with  $\alpha > 0$  constant. Specifically, in [3][12], this differential equation can be solved with the following three conditions for the function  $f(v)$  and  $g(v)$ .

*Condition 1.*  $f(\cdot)$  is piecewise smooth, monotone increasing, odd, with  $\lim_{v \rightarrow \infty} f'(v)$  finite;

*Condition 2.*  $g(\cdot)$  is piecewise continuous, even, with

$$\lim_{v \rightarrow \infty} g(v) = \lim_{v \rightarrow \infty} f'(v) \quad (4)$$

*Condition 3.*  $f'(v) > g(v) > \alpha e^{\alpha v} \int_v^\infty |f'(\zeta) - g(\zeta)| e^{-\alpha \zeta} d\zeta$  for all  $v > 0$ .

Then, the solution of the equation (3) can be re-expressed as

$$w = f(v) + [w_0 - f(v_0)]e^{-\alpha(v-v_0)sgn(\dot{v})} + e^{-\alpha v sgn(\dot{v})} \int_{v_0}^v [g(\zeta) - f'(\zeta)]e^{\alpha \zeta sgn(\dot{v})} d\zeta \quad (5)$$

where  $(v_0, w_0)$  is the initial value.

For convenience, the equation (5) can be re-expressed as

$$w = f(v) + \varphi(v) \quad (6)$$

where

$$\varphi(v) = [w_0 - f(v_0)]e^{-\alpha(v-v_0)sgn(\dot{v})} + e^{-\alpha v sgn(\dot{v})} \int_{v_0}^v [g(\zeta) - f'(\zeta)]e^{\alpha \zeta sgn(\dot{v})} d\zeta$$

*Property 1.* The nonlinear function  $\varphi(v)$  is bounded, and  $\varphi(v) \rightarrow 0$  when  $v \rightarrow \infty$ .

The boundedness of the function  $\varphi(v)$  is an important property for controller design. The proof will be shown as follows briefly [4].

When  $v$  is increasing,  $\text{sgn}(\dot{v}) = 1$ , and the solution  $w_L$  of (3) that starts at  $(v_0, w_0)$  is

$$\begin{aligned} w &= w_L(v) = w_L(v; v_0, w_0) \\ &= f(v) + [w_0 - f(v_0)]e^{-\alpha(v-v_0)} - e^{-\alpha v} \int_{v_0}^v [f'(\zeta) - g(\zeta)]e^{\alpha\zeta} d\zeta, \quad v \geq v_0 \end{aligned} \quad (7)$$

When  $v$  is decreasing,  $\text{sgn}(\dot{v}) = -1$ , and the solution of (3) issuing from  $(v_0, w_0)$  is

$$\begin{aligned} w &= w_U(v) = w_U(v; v_0, w_0) \\ &= f(v) + [w_0 - f(v_0)]e^{-\alpha(v_0-v)} - e^{\alpha v} \int_v^{v_0} [f'(\zeta) - g(\zeta)]e^{\alpha\zeta} d\zeta, \quad v \leq v_0 \end{aligned} \quad (8)$$

Obviously, according to Condition 3, it can be obtained that

$$\lim_{v \rightarrow +\infty} [w_L(v; v_0, w_0) - f(v)] = 0 \quad (9)$$

$$\lim_{v \rightarrow -\infty} [w_U(v; v_0, w_0) - f(v)] = 0 \quad (10)$$

Therefore, it can be concluded that the term  $\varphi(v)$  is bounded and when  $v \rightarrow \infty$ ,  $\varphi(v) \rightarrow 0$ .

*Remark 1.* The advantage of the Duhem model is that various hysteresis shapes can be expressed by choosing different functions  $f(v)$  and  $g(v)$ , which satisfy three conditions. When  $f(v)$  and  $g(v)$  are selected as a linear function and a constant, the Duhem model can also describe the backlash-like nonlinearities [18]. The motivation of using the Duhem model in this paper is to conduct the suitable control approach for more general hysteresis input. Especially, the control scheme also is applicable to the nonlinear systems preceded by Backlash-like nonlinearities [18].

## 4 Controller Design

In this section, the control scheme for the nonlinear systems with unknown hysteresis input represented by the Duhem model is discussed. The plant defined in (1) with the input  $w(t)$  described by Duhem model (3) can be expressed as

$$x^{(n)}(t) + \sum_{i=1}^k a_i Y_i(x(t), \dot{x}(t), \dots, x^{(n-1)}(t)) = bf(v(t)) + b\varphi(v(t)) \quad (11)$$

and the plant can also be re-expressed as

$$\begin{aligned} \dot{x}_i &= x_{i+1}, i = 1, \dots, n-1 \\ \dot{x}_n &= -\sum_{i=1}^k a_i Y_i(\mathbf{x}) + bf(v(t)) + b\varphi(v(t)) \end{aligned} \quad (12)$$

where  $x_1 = x, x_2 = \dot{x}, \dots, x_n = x^{(n-1)}$  and  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ ,

*Remark 2.* When the control input is a linear function, the control schemes have been discussed in [18] [25]. For the controlled plant (11), the challenge for controller design is the existence of the function  $f(v)$ , which is an unknown nonlinear function. Therefore, a special new treatment is required to deal with  $f(v)$ , which constitutes the main contribution of the paper.

In order to design the controller for the plant (11), the following assumptions about the plant and the Duhem model are required.

*Assumption 2.* There exist known positive constants  $b_{min}$  and  $D$  such that control gain satisfies  $b_{min} \leq b$  and the nonlinear item  $b\varphi(v)$  satisfies  $|b\varphi(v)| \leq D$ .

*Assumption 3.* There exist constants  $a_{imax}, i = 1, \dots, k$  such that  $a_i \leq a_{imax}$ .

*Assumption 4.* The nonlinear input function of Duhem model  $f(v)$  satisfies the following inequality:

$$vf(v) \geq \alpha_1 v^2 \quad (13)$$

where  $\alpha_1$  is known and satisfies  $\alpha_1 > 0$ .

*Remark 3.* As introduced in Section 3,  $\varphi(v)$  is bounded and  $\varphi(v) \rightarrow 0$  when  $v \rightarrow \infty$ . It is reasonable to assume the bound of the  $|b\varphi(v)|$  in Assumption 2 based on the open-loop experiment data. Assumption 3 is common for the nonlinear system control design. Numerous hysteresis shapes in the smart materials based actuators, such as magnetostrictive actuators, can satisfy Condition 3 of Duhem model easily, therefore, it is reasonable to assume  $f(v)$  satisfying this inequality in Assumption 4 when  $f(v)$  is an odd function.

In order to present the robust control law, a tracking error vector  $e$  is defined as

$$\mathbf{e} = \mathbf{x} - \mathbf{x}_d \quad (14)$$

and a filtered tracking error is given as

$$e_s = \left( \frac{d}{dt} + \lambda \right)^{(n-1)} e_1 \quad \lambda > 0 \quad (15)$$

where  $e_1 = x_1 - x_d$ .

Also,  $e_s$  can be re-written as  $e_s = [\Lambda^T \ 1] e$  with  $\Lambda = [\lambda^{(n-1)}, (n-1)\lambda^{(n-2)}, \dots, (n-1)\lambda]^T$ .

It has been shown in [11] and [16] that the filtered tracking error has the following properties.

1). The equation  $e_s = 0$  defines a boundary layer  $S_\varepsilon$ , satisfying  $S_\varepsilon = \{\mathbf{e} \in \mathbf{R}^n \mid |e_s| \leq \varepsilon\}$ , where  $\varepsilon$  is a strictly positive constant.

2). The tracking error vector will exponentially converge to an  $n$ -dimensional box defined according to  $\{|e_i| \leq \zeta_i \lambda^{i-n+1} \varepsilon, i = 0, 1, \dots, n-1\}$  with  $\zeta_i$  satisfying

$$\zeta_i = \begin{cases} 1, & \text{for } i = 0 \\ 1 + \sum_{j=0}^{i-1} \binom{i}{j} \zeta_j & \text{for } i = 1, 2, \dots, n-1. \end{cases} \quad (16)$$

In this paper, a tuning error is introduced as

$$e_{\varepsilon s} = e_s - \varepsilon \text{sat}\left(\frac{e_s}{\varepsilon}\right) \quad (17)$$

where  $\text{sat}(\cdot)$  is the saturation function.

For the plant preceded by the Duhem hysteresis model and subjected to the above assumptions, the following robust controller is conducted

$$v(t) = -\text{sgn}(e_{\varepsilon s})\mu(x, t) \quad (18)$$

where

$$\mu(x, t) = \frac{\beta}{b_{\min} \alpha_1} \left( \sum_{i=1}^k |a_{i\max} Y_i(\mathbf{x}(t))| + D + |\nu| \right) \quad (19)$$

with  $\nu = -x_d^{(n)} + [0 \ A^T]e$ , and  $\beta > 1$ .

The stability of the closed-loop system is established in the following theorem:

**Theorem 1.** *For the closed-loop system consisting of the plant [1] preceded by unknown hysteresis described by Duhem model in [3], subjected to Assumptions 1-4, the robust controller defined in [18] can guarantee the global boundedness of the closed-loop system and the output tracking with a desired precision.*

*Proof:* To establish the global boundedness, the following Lyapunov function candidate is chosen as:

$$V(t) = \frac{1}{2} e_{\varepsilon s}^2 \quad (20)$$

Differentiating  $V(t)$  with respect to time  $t$  leads to

$$\dot{V}(t) = e_{\varepsilon s} \dot{e}_{\varepsilon s} \quad (21)$$

When  $|e_s| \leq \varepsilon$ , the derivative  $\dot{V}(t) = 0$  based on the definition of  $e_{\varepsilon s}$ . When  $|e_s| > \varepsilon$ , the time derivative of the filtered error is

$$\begin{aligned} \dot{e}_s &= - \sum_{i=1}^k a_i Y_i(\mathbf{x}(t)) + \nu + bw(t) \\ &= - \sum_{i=1}^k a_i Y_i(\mathbf{x}(t)) + \nu + bf(v(t)) + b\varphi(v(t)) \end{aligned} \quad (22)$$

When  $|e_s| > \varepsilon$ , we know the fact that  $e_{\varepsilon s}\dot{e}_{\varepsilon s} = e_{\varepsilon s}\dot{e}_s$ . Using (22), it has

$$e_{\varepsilon s}\dot{e}_s = e_{\varepsilon s}\left(-\sum_{i=1}^k a_i Y_i(\mathbf{x}(t)) + \nu + bf(v(t)) + b\varphi(v(t))\right) \quad (23)$$

According to Eq. (23), the following inequality can be obtained

$$e_{\varepsilon s}\dot{e}_s \leq |e_{\varepsilon s}|\left(\sum_{i=1}^k |a_i Y_i(\mathbf{x}(t))| + b\operatorname{sgn}(e_{\varepsilon s})f(v(t)) + D + |\nu|\right) \quad (24)$$

When  $e_{\varepsilon s} > 0$ , the controller defined in (18) can be expressed as  $v = -\mu(x, t)$ , then the inequality (13) can be deduced as

$$vf(v) = -\mu(x, t)f(v) \geq \alpha_1 v^2 = \alpha_1 \mu^2(x, t)$$

According to the definition of  $\mu$  in (19), it is obvious that  $\mu \geq 0$ , then it can be obtained that  $f(v) \leq -\alpha_1 \mu(x, t)$ , when  $e_{\varepsilon s} > 0$ .

Similarly, when  $e_{\varepsilon s} < 0$ , we have  $v = \mu(x, t)$ , then the inequality (13) can be deduced as

$$vf(v) = \mu(x, t)f(v) \geq \alpha_1 v^2 = \alpha_1 \mu^2(x, t) \quad (25)$$

and it can also be obtained that  $-f(v) \leq -\alpha_1 \mu(x, t)$ , when  $e_{\varepsilon s} < 0$ .

Then, it can be concluded that

$$\operatorname{sgn}(e_{\varepsilon s})f(v) \leq -\alpha_1 \mu(x, t) \quad (26)$$

Using the conclusion in Eq. (26), the inequality (24) can be deduced as

$$e_{\varepsilon s}\dot{e}_s \leq |e_{\varepsilon s}|\left(\sum_{i=1}^k |a_i Y_i(\mathbf{x}(t))| - b\alpha_1 \mu(x, t) + D + |\nu|\right) \quad (27)$$

According to the definition of  $\mu(x, t)$ , the inequality (27) can re-expressed as

$$\begin{aligned} e_{\varepsilon s}\dot{e}_s &\leq |e_{\varepsilon s}|\left(\sum_{i=1}^k |a_i Y_i(\mathbf{x}(t))| - b\alpha_1 \mu(x, t) + D + |\nu|\right) \\ &\leq |e_{\varepsilon s}|\left(\sum_{i=1}^k |a_i Y_i(\mathbf{x}(t))| + D + |\nu| - b\alpha_1 \frac{\beta}{b_{\min}\alpha_1} \left(\sum_{i=1}^k |a_{i\max} Y_i(\mathbf{x}(t))| + D + |\nu|\right)\right) \end{aligned}$$

then we have

$$\begin{aligned} e_{\varepsilon s}\dot{e}_s &\leq |e_{\varepsilon s}|\left(\sum_{i=1}^k |a_i Y_i(\mathbf{x}(t))| + D + |\nu| - b\frac{\beta}{b_{\min}} \left(\sum_{i=1}^k |a_{i\max} Y_i(\mathbf{x}(t))| + D + |\nu|\right)\right) \\ &\leq |e_{\varepsilon s}|\left(\sum_{i=1}^k |a_i Y_i(\mathbf{x}(t))| + D + |\nu| - \beta \left(\sum_{i=1}^k |a_{i\max} Y_i(\mathbf{x}(t))| + D + |\nu|\right)\right) \\ &\leq |e_{\varepsilon s}|\left((1 - \beta)(D + |\nu|) + |e_{\varepsilon s}|\left(\sum_{i=1}^k |a_i Y_i(\mathbf{x}(t))| - \beta \sum_{i=1}^k |a_{i\max} Y_i(\mathbf{x}(t))|\right)\right) \end{aligned}$$

It is obvious that  $\sum_{i=1}^k |a_i Y_i(\mathbf{x}(t))| \leq \sum_{i=1}^k |a_{imax} Y_i(\mathbf{x}(t))|$  by utilizing Assumption 3, then the above inequality can be deduced as

$$e_{\varepsilon s} \dot{e}_s \leq |e_{\varepsilon s}|(1 - \beta)(D + |\nu| + \sum_{i=1}^k |a_{imax} Y_i(\mathbf{x}(t))|)$$

Since  $\beta > 1$ , it is obvious that  $e_{\varepsilon s} \dot{e}_s \leq 0$ . The proposed robust control approach ensures that  $\dot{V} \leq 0$ . Therefore, it can be concluded that  $V$  defined in (20) is nonincreasing, and  $e_{\varepsilon s}$  is bounded. Since  $\dot{V} \leq 0$ , it can be proved that  $\dot{V}(t) \rightarrow 0$  for all  $t$ . Therefore, it can be shown that  $e_{\varepsilon s} \rightarrow 0$ , when  $t \rightarrow \infty$ , then  $\mathbf{e}$  will exponentially converge to an  $n$ -dimensional box defined hereinbefore.

*Remark 4.* the main purpose of this paper is to explore the possibility to combine the robust control approach with Duhem model. The future work will relax the assumptions defined in this paper to make it easier to apply to the industrial systems.

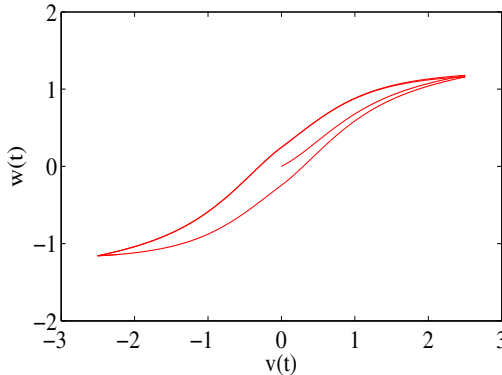
### 5 Simulation

To illustrate the proposed robust control algorithms, we consider a nonlinear system in the form (II) given by

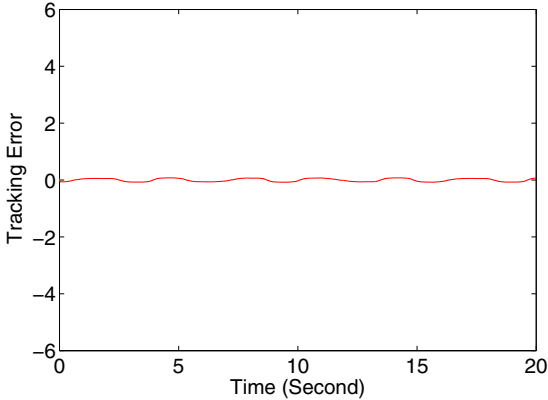
$$\dot{x} = a \frac{1 - e^{-x(t)}}{1 + e^{-x(t)}} + bw(t) \tag{28}$$

where  $w(t)$  represents the output of the hysteresis described by the Duhem model. Therein, the nonlinear functions  $f(v)$  and  $g(v)$  are chosen as

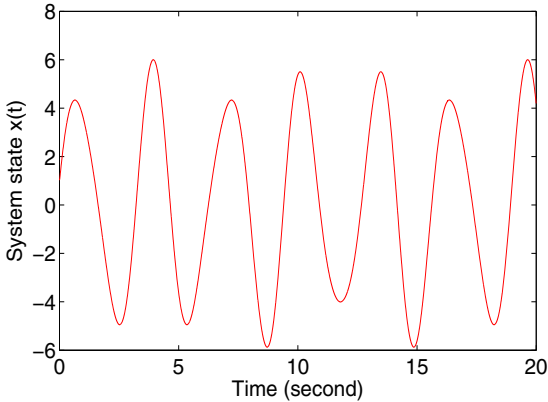
$$\begin{aligned} f(v) &= \tanh(v) + 0.1v \\ g(v) &= f'(v)(1 - 0.58e^{-|v|}) \end{aligned} \tag{29}$$



**Fig. 1.** Hysteresis curve given by Duhem model



**Fig. 2.** Tracking error of the state with Duhem hysteresis

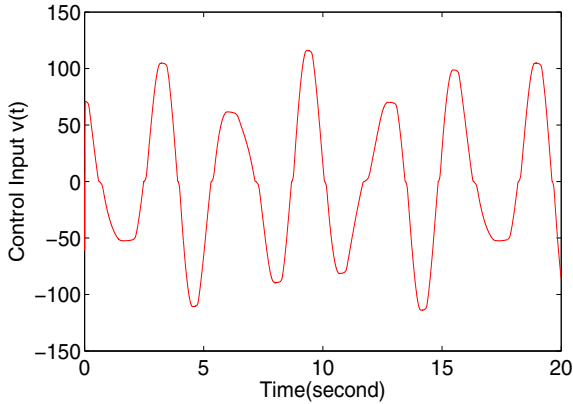


**Fig. 3.** The system state  $x(t)$  with Duhem hysteresis

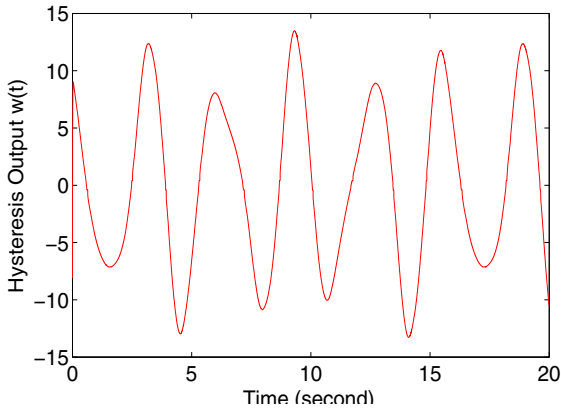
In this case, we know that  $f'(v) > g(v)$  for all finite  $v$  and  $g(\infty) = f'(\infty) = 0.1$ . The input signal is chosen as  $v(t) = 2.5 \sin(2.3t)$ , and the hysteresis shape constructed by Duhem model is shown in Fig. 1. From Fig. 1, the strong hysteresis preceding the controlled plant has multi-values and saturation properties, which can describe the hysteresis with saturation property, such as magnetostrictive actuators.

Also, the following inequality can be obtained that  $vf(v) = (\tanh(v)+0.1v)v = \tanh(v)v + 0.1v^2 \geq 0.1v^2$  satisfying Assumption 4. The plant parameter values are selected as  $a = 1$  and  $b = 1$  and the control design parameters are chosen as  $\alpha_1 = 0.1$ ,  $b_{min} = 1$ ,  $\beta = 10$ ,  $D = 3$ ,  $\alpha_{1max} = 1$  and  $\varepsilon = 0.005$ . The initial state is  $x(0) = 1.05$ .





**Fig. 4.** Control signal  $v(t)$  acting as the input of Duhem hysteresis



**Fig. 5.** Hysteresis output  $w(t)$  acting as the input of system

In order to illustrate the tracking performance with the designed controller, choosing the desired trajectory as  $x_d(t) = 5\sin(2t) + \cos(3.2t)$ , the simulation results are shown in Figs. 2 - 5 in the time spend of 1 and 20s. Figs. 2 and 3 show the tracking error and the state trajectory. The control signal is shown in Fig. 4 and the output of the hysteresis is shown in Fig. 5. It should be mentioned that it is difficult to compare the control performance with considering hysteresis or without considering hysteresis as [24], since the hysteresis represented by a differential equation can not be divided into linear part and hysteresis part easily. The simulation results show that the proposed control scheme can overcome the effects of the hysteresis and keep the tracking performance of the closed-loop system, which can justify the effectiveness of the designed controller.

## 6 Conclusion

In this paper, the unknown hysteresis nonlinearities are presented by the a differential equation-based model, Duhem hysteresis model, and a corresponding robust control architecture is proposed for a class of preceded hysteretic nonlinear systems. The main contribution of this paper is the properties of Duhem model are explored, which makes it possible to fuse this hysteresis model with the robust control approach. The developed control law ensures the global boundedness of the closed-loop system and achieves tracking with a desired precision. Simulation results has confirmed the effectiveness of the proposed control approach.

## Acknowledgement

The work was partially supported by “the Fundamental Research Funds for the Central Universities, SCUT(2009ZZ0005)”.

## References

1. Bessa, W.M.: Some remarks on the boundedness and convergence properties of smooth sliding mode controllers. *International Journal of Automation and Computing* 6(2), 154–158 (2009)
2. Brokate, M., Sprekels, J.: *Hysteresis and Phase Transitions*. Springer, Berlin (1996)
3. Chen, X., Hisayama, T., Su, C.-Y.: Adaptive control for continuous-time systems preceded by hysteresis. *IEEE Trans. Automat. Contr.* 53(4), 1019–1025 (2008)
4. Chua, L.O., Stromsmoe, K.A.: Mathematical model for dynamic hysteresis loops. *Int. J. Engng. Sci.* 9, 435–450 (1971)
5. Coleman, B.D., Hodgdon, M.L.: On a class of constitutive relations for ferromagnetic hysteresis. *Arch. Rational Mech. Anal.*, 375–396 (1987)
6. Duhem, P.: Die dauernden Aenderungen und die Thermodynamik. I, *Z. Phys. Chem.* 22, 543–589 (1897)
7. Feng, Y., Hu, Y.-M., Rabbath, C.A., Su, C.-Y.: Robust adaptive control for a class of perturbed strict-feedback nonlinear systems with unknown Prandtl-Ishlinskii hysteresis. *International Journal of Control* 81(11), 1699–1708 (2008)
8. Krasnosel’skii, M.A., Pokrovskii, A.V.: *Systems with Hysteresis*. Springer, New York (1989)
9. Klein, O., Krejci, P.: Outwards pointing hysteresis operators and asymptotic behaviour of evolution equations. *Nonlinear Analysis: Real World Applications* 4(5), 755–785 (2003)
10. Krejci, P., Kuhnen, K.: Inverse control of systems with hysteresis and creep. *IEE Proc.-Control Theory and Applications* 148(3), 185–192 (2001)
11. Liu, S., Huang, T., Yen, J.: Tracking control of shape-memory-alloy actuators based on self-sensing feedback and inverse hysteresis compensation. *Sensor* 10, 112–127 (2010)
12. Macki, J.W., Nistri, P., Zecca, P.: Mathematical models for hysteresis. *SIAM Review* 35(1), 94–123 (1993)
13. Mayergoyz, I.D.: *Mathematical Models of Hysteresis*. Springer, New York (1991)

14. Oh, J., Bernstein, D.S.: Semilinear Duhem model for rate-independent and rate-dependent hysteresis. *IEEE Trans, Automat. Contr.* 50(5), 631–645 (2005)
15. Oh, J., Bernstein, D.S.: Piecewise linear identification for the rate-independent and rate-dependent Duhem hysteresis models. *IEEE Trans. Automat. Contr.* 52(3), 576–582 (2007)
16. Slotine, J.J.E., Li, W.: *Applied Nonlinear Control*. China Machine Press (2004)
17. Shyu, K.K., Liu, W.J., Hsu, K.C.: Decentralised variable structure control of uncertain large-scale systems containing a dead-zone. *IEE Proc. Control Theory Appl.* 150(5), 467–475 (2003)
18. Su, C.-Y., Stepanenko, Y., Svoboda, J., Leung, T.P.: Robust adaptive control of a class of nonlinear systems with unknown backlash-like hysteresis. *IEEE Trans. Automat. Contr.* 45(12), 2427–2432 (2000)
19. Feng, Y., Hu, Y.-M., Rabbath, C.A., Su, C.-Y.: Robust adaptive control for a class of perturbed strict-feedback nonlinear systems with unknown Prandtl-Ishlinskii hysteresis. *International Journal of Control* 81(11), 1699–1708 (2008)
20. Tan, X., Baras, J.S.: Modeling and Control of Hysteresis in Magnetostrictive Actuators. *Automatica* 40(9), 1469–1480 (2004)
21. Tan, X., Baras, J.S.: Adaptive identification and control of hysteresis in smart materials. *IEEE Trans, Automat. Contr.* 50(16), 827–839 (2005)
22. Tao, G., Kokotović, P.V.: Adaptive control of plants with unknown hysteresis. *IEEE Trans. Automat. Contr.* 40, 200–212 (1995)
23. Visintin, A.: *Differential Models of Hysteresis*. Springer, New York (1994)
24. Wang, Q., Su, C.-Y.: Robust adaptive control of a class of nonlinear systems including actuator hysteresis with Prandtl-Ishlinskii presentations. *Automatica* 42(5), 859–867 (2006)
25. Zhou, J., Wen, C.-Y., Zhang, Y.: Adaptive backstepping control of a class of uncertain nonlinear systems with unknown backlash-like hysteresis. *IEEE Trans. Automat. Contr.* 49(10), 1751–1757 (2004)

# A Generalized Play Operator for Modeling and Compensation of Hysteresis Nonlinearities

Mohammad Al Janaideh<sup>1</sup>, Omar Aljanaideh<sup>2</sup>, and Subhash Rakheja<sup>2</sup>

<sup>1</sup> Department of Mechatronics Engineering, Amman,  
The University of Jordan, Amman, Jordan  
aljanaideh@gmail.com

<sup>2</sup> Department of Mechanical and Industrial Engineering,  
Concordia University, Montreal, Canada  
omaryanni@gmail.com, rakheja@alcor.concordia.ca

**Abstract.** Smart actuators, such as shape memory alloy and magnetostrictive actuators, exhibit saturation nonlinearity and hysteresis that may be symmetric or asymmetric. The Prandtl–Ishlinskii model employing play operators has been used to describe the hysteresis properties of smart actuators that are symmetric in nature. In this paper, the application of a generalized play operator capable of characterizing symmetric as well as asymmetric hysteresis properties with output saturation is explored in formulating a generalized Prandtl–Ishlinskii model. Since the proposed generalized model is a mere extension of the analytically invertible Prandtl–Ishlinskii model, an inverse of the generalized model is formulated using the inverse of the Prandtl–Ishlinskii model together with those of the envelope functions of the generalized play operator. The effectiveness of the inverse of the generalized model in compensating for the hysteresis effects is subsequently investigated through simulations and experiments performed on a piezoceramic actuator. The simulation results suggest that the inverse of the generalized Prandtl–Ishlinskii model can be conveniently applied as a feedforward controller to compensate for hysteresis nonlinearities.

**Keywords:** Prandtl–Ishlinskii, play operator, hysteresis.

## 1 Introduction

Smart actuators such as magnetostrictive, shape memory alloys, and piezoceramic actuators are widely used in micropositioning applications. Such actuators generally exhibit strong hysteresis effects in their output responses, which may cause inaccuracies and oscillations in the closed-loop system responses [1]. Such oscillations could be particularly detrimental in applications involving positioning measurement and control. A number of models have been reported to describe hysteresis properties of smart actuators to achieve accurate micropositioning control through compensation of the hysteresis effects. These models include Preisach model, Krasnosel’skii–Pokrovskii model, and Prandtl–Ishlinskii model [2].

In the Preisach formulation, hysteresis is modeled as a cumulative effect (density function) of all possible relay elements (relay operators), which are parameterized by

a pair of threshold variable. Two other hysteresis models, Krasnosel'skii-Pokrovskii model and Prandtl-Ishlinskii model, have been derived from the Preisach model. The Krasnosel'skii-Pokrovskii model is based on an integral of a density function and Krasnosel'skii-Pokrovskii operators, defined by two functions that are bounded by two piecewise Lipschitz continuous functions [1]. The Prandtl-Ishlinskii model is a superposition of elementary play or stop operators, which are parameterized by a single threshold variable. This model is defined in terms of an integral of play or stop operator with a density function determining the shape of the hysteresis loop. The Prandtl-Ishlinskii model offers advantage over the Preisach model, since its inverse can be computed analytically, which makes it extremely attractive for real-time control applications, particularly for real-time compensation of hysteresis contributions [2]. The Prandtl-Ishlinskii model has been applied to characterize symmetric and rate-independent hysteresis properties of materials and smart actuators. The model can yield considerable errors when an asymmetry exists in the hysteresis loops, such as those observed in the output-input properties of the shape memory alloy and magnetostrictive actuators, or when the output-input relations are dependent on the rate of the applied input [4].

In this paper, a generalized Prandtl-Ishlinskii model is formulated to describe the hysteresis properties of smart actuators. For this purpose a generalized play hysteresis operator is formulated and integrated to the Prandtl-Ishlinskii model. Since the resulting generalized model is a mere extension of the analytically invertible Prandtl-Ishlinskii model, an analytical inverse of the generalized Prandtl-Ishlinskii model is obtained using the inverse for both the Prandtl-Ishlinskii model and the envelope functions. However, this requires that the envelope functions be invertible.

## 2 Prandtl-Ishlinskii Model

The Prandtl-Ishlinskii model utilizes the play operator and positive weights to describe the input-output hysteresis nonlinearities in actuators and materials [1]. The play operator is constructed with the current input, the previous output, and the threshold  $r$ , which is a positive constant. This constant handles the hysteresis nonlinearity of the play operator and the Prandtl-Ishlinskii hysteresis model. The play operator  $\Xi_r$  has also been described by the motion of a piston within a cylinder of length  $2r$  where the instantaneous position of center of the piston is represented by the coordinate  $u$  and that of the cylinder position by  $w$  [2]. Analytically, for any  $t \in [0, T]$  the input  $u(t) \in C[0, T]$ , where  $C[0, T]$  is the space of continuous function, the output of the play operator is described as :

$$w(t) = \Xi_r[u](t) = \max\{ u(t) - r, \min\{ u(t) + r, \Xi_r[u](0) \} \} \tag{1}$$

where

$$w(0) = \Xi_r[u](0) = \max\{ u(0) - r, \min\{ u(0) + r, 0 \} \} \tag{2}$$

In the above formulation,  $r$  refers to the threshold value, which is the magnitude of increasing or decreasing input  $u(t)$  corresponding to zero output  $w(t)$ . The play operator is continuous, rate-independent, and symmetric hysteresis operator. The argument

of the operator is written in square brackets to indicate its functional dependence, since it maps a function to another function. It has been shown that for any input  $C[0, T]$ , the play operator is Lipschitz continuous and monotone [3].

Since a finite number of hysteresis play operators with the weights  $\theta_j$  are adequate to model for hysteresis nonlinearities in practical applications [2], the output of the Prandtl-Ishlinskii model can be expressed with  $N$  play hysteresis operators as:

$$y(t) = \Omega[u](t) = \sum_{j=0}^N \theta_j \Xi_{r_j}[u](t) \quad (3)$$

For any input  $u(t) \in C[0, T]$ , the output of the Prandtl-Ishlinskii model is  $C[0, T]$ . The Prandtl-Ishlinskii model has been applied to characterize hysteresis effects of piezoceramic actuators [3]. The analytical inverse of the Prandtl-Ishlinskii model has been applied to compensate for hysteresis nonlinearities of piezoceramics actuators [3].

The use of the Prandtl-Ishlinskii model to characterize asymmetric hysteresis loops may yield considerable characterization errors. Consequently, the use of the inverse Prandtl-Ishlinskii model to compensate for asymmetric and saturated hysteresis nonlinearities may lead to considerable errors in the output compensation [4].

### 3 Generalized Prandtl-Ishlinskii Model

Generalized play operator is proposed in this paper to relax the symmetric property of the play hysteresis operator and to develop a generalized Prandtl-Ishlinskii model for characterizing asymmetric and saturated hysteresis nonlinearities of smart actuators. The inverse of the proposed model can be also obtained to compensate for hysteresis nonlinearities.

#### 3.1 Generalized Play Operator

In the generalized operator  $\Phi_r$ , an increase in input  $u$  causes the output of the generalized operator  $m$  to increase along the curve  $\Gamma_f$ , while a decrease in input  $u$  causes the output  $m$  to decrease along the curve  $\Gamma_g$ , resulting in an asymmetric loop. The generalized play operator is constructed with the following condition:

$$\Gamma_f(t) \leq \Gamma_g(t) \quad (4)$$

where the  $\Gamma_f$  and  $\Gamma_g$  curves are defined as:

$$\Gamma_f(t) = f\{u(t)\} - r \quad (5)$$

$$\Gamma_g(t) = g\{u(t)\} + r \quad (6)$$

The functions  $f$  and  $g$  are strictly increasing continuous odd function. Analytically, the output of the generalized play operator for any input  $u(t) \in C[0, T]$  is defined as:

$$m(t) = \Phi_r[u](t) = \max\{ \Gamma_f(t), \min\{ \Gamma_g(t), \Phi_r[u](0) \} \} \quad (7)$$

where

$$m(0) = \Phi_r[u](0) = \max\{ \Gamma_f(0), \min\{ \Gamma_g(0), 0 \} \} \quad (8)$$

Unlike the play operator, the generalized play operator constructed with invertible envelope functions yields zero outputs,  $m(t) = 0$ , at two different values,  $\alpha$  and  $\beta$ , of the increasing and decreasing inputs  $u(t)$ . The values of  $\alpha$  and  $\beta$  are expressed as:

$$\alpha = f^{-1}(r) \tag{9}$$

$$\beta = g^{-1}(-r) \tag{10}$$

The difference in the magnitudes of  $\alpha$  and  $\beta$  allows for describing asymmetric hysteresis loops.

### 3.2 Input-Output Relationship of the Generalized Prandtl-Ishlinskii Model

The generalized Prandtl-Ishlinskii model  $\Delta$  is subsequently formulated upon integrating the generalized play operator  $\Phi_r$  and the weights  $\theta_j$  in order to model the asymmetric and saturated hysteresis nonlinearities. The output of the generalized Prandtl-Ishlinskii model  $\Delta[u](t)$ , which maps  $C[0, T]$  into  $C[0, T]$ , is analytically expressed using  $N$  generalized play hysteresis operators as:

$$m(t) = \Delta[u](t) = \sum_{j=0}^N \theta_j \Phi_{r_j}[u](t) \tag{11}$$

The generalized Prandtl-Ishlinskii model can be used to model asymmetric hysteresis nonlinearities in piezoceramic actuators as well as shape memory alloy actuators.

## 4 Hysteresis Modeling

The parameters of the generalized play hysteresis operator and the weights need to be defined on the basis of known characteristics of smart actuators. In this study, the experimental data reported for a shape memory alloy actuator [5] used to identify the model parameters. The generalized play hysteresis operator is constructed with envelope functions to characterize the observed nonlinearities in the measured data. The hyperbolic tangent envelope functions are chosen so as to describe output saturation under certain input, such that:

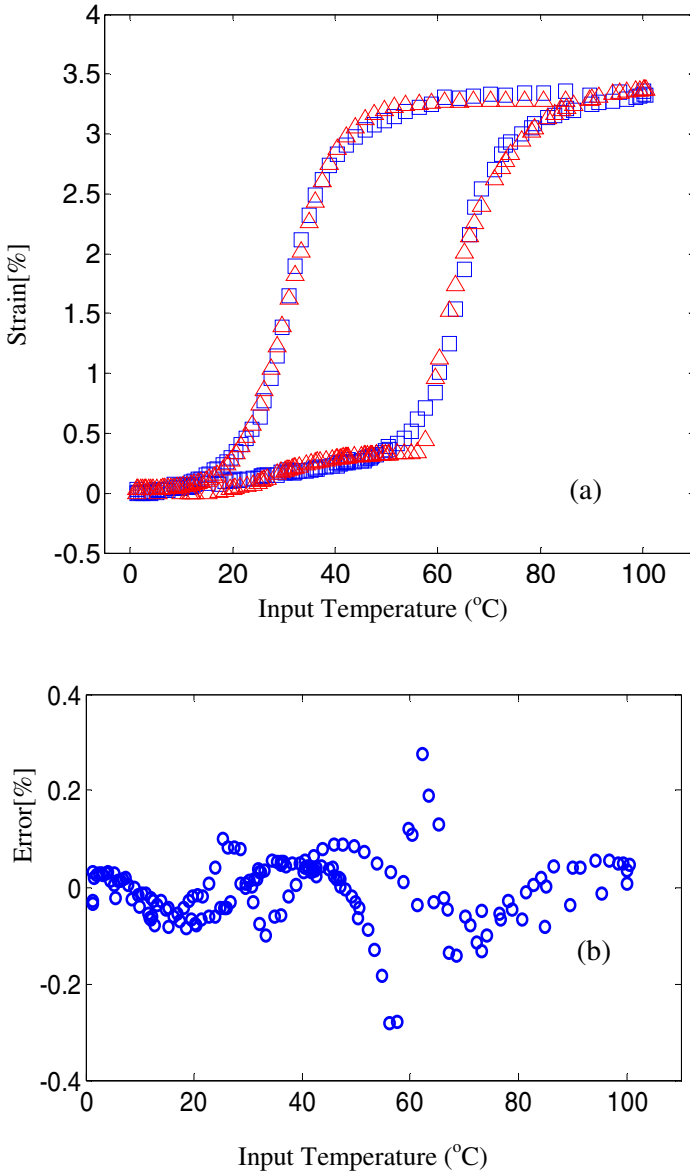
$$\begin{aligned} f(u) &= a_0 \tanh(a_1 u + a_2) + a_3 \\ g(u) &= b_0 \tanh(b_1 u + b_2) + b_3 \end{aligned} \tag{12}$$

The generalized Prandtl-Ishlinskii model is formulated using the following threshold and weights values :

$$r_j = \delta \ j \tag{13}$$

$$\theta_j = \rho \ e^{-\tau \ r_j} \tag{14}$$

where  $a_0, a_1, a_2, a_3, b_0, b_1, b_2, b_3, \rho > 0, \tau$ , and  $\delta > 0$  are constants, which are identified from the measured data. It should be mentioned that the proposed weight and threshold functions of the generalized play operator are not unique; they depend upon the nature of hysteresis of particular material or actuator. The model parameters were identified through minimization of an error sum-squared.



**Fig. 1.** (a) Comparison of responses of the generalized Prandtl-Ishlinskii model with measured data of NiTi Shape Memory Alloy under constant tensile stress of 50 MPa (Triangular - the measured strain percent; Square -The output of the generalized Prandtl-Ishlinskii model), (b) the error of the generalized Prandtl-Ishlinskii model.



$$Q(X) = \sum_{l=1}^L (\eta(l) - \eta_m(l))^2 \tag{15}$$

where  $\eta$  is the displacement response of the generalized Prandtl-Ishlinskii model to a given input,  $\eta_m$  is measured displacement response of the actuator to the same input,  $L$  is the number of data points considered,  $Q$  is the error function and  $X$  is the parameters vector, given by:  $\{X\} = \{a_o, a_1, a_2, a_3, b_o, b_1, b_2, b_3, \tau, \rho, \text{ and } \delta\}$ . The error minimization problem was solved using the MATLAB optimization toolbox, subject to following constraints:  $\{a_o, a_1, b_o, b_1, \rho, \tau, \delta\} > 0$ .

The solutions of the minimization problem were attained using measured data for both actuators for different starting parameter vectors, which generally converged to similar solutions. The effectiveness of the generalized model in predicting the saturated hysteresis responses of the shape memory alloy actuator can also be seen from comparisons of the displacement responses in Fig.1(a). The figure represents the comparison of the model responses with the measured data. The error responses of the model is shown Fig.1(b). The figure shows that the peak error in the output responses of the generalized Prandtl-Ishlinskii model is approximately 0.3 %.

## 5 The Inverse Generalized Prandtl-Ishlinskii Model

The concept of the open loop control system used in this paper to obtain identity mapping between the desired input  $u(t)$  and the compensation output  $v(t)$  such that :  $u(t) = v(t)$ . In this section, the inverse generalized Prandtl-Ishlinskii model is presented. Suppose that the inverse of the envelope functions of the generalized play operator exist ( $f^{-1}, g^{-1}: \mathbb{R} \rightarrow \mathbb{R}$ ) and the output of the inverse envelope functions are Lipchitz continuous, the exact inverse of the generalized Prandtl-Ishlinskii model can be obtained analytically. The output of generalized Prandtl-Ishlinskii model  $\Delta^{-1}[u](t)$  can be expressed as:

$$\Delta^{-1}[u](t) = \begin{cases} f^{-1} \circ \Omega_+^{-1}[u](t) & \text{for } \dot{u}(t) \geq 0 \\ g^{-1} \circ \Omega_-^{-1}[u](t) & \text{for } \dot{u}(t) \leq 0 \end{cases} \tag{16}$$

where  $\Omega [u](t)$  is the output of the inverse Prandtl-Ishlinskii model when the desired input increases and decreases, respectively. Analytically:

$$\Omega^{-1}[u](t) = \begin{cases} \Omega_+^{-1}[u](t) & \text{for } \dot{u}(t) \geq 0 \\ \Omega_-^{-1}[u](t) & \text{for } \dot{u}(t) \leq 0 \end{cases} \tag{17}$$

Since the inverse of the Prandtl-Ishlinskii model is a Prandtl-Ishlinskii model, the output of the inverse model can be expressed as [3]:

$$\Omega^{-1}[u](t) = \hat{\theta}_0 u(t) + \sum_{j=1}^N \hat{\theta}_j \Xi_s[u](t) \tag{18}$$

where the weights of the inverse model are defined for  $j = 1, 2, \dots, N$  as [3]:

$$\hat{\theta}_0 = \frac{1}{\theta_0}$$

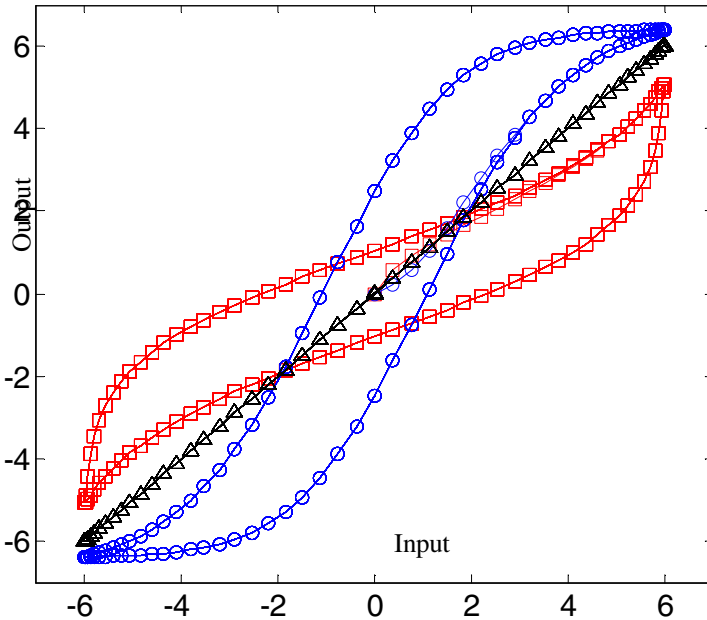
$$\hat{\theta}_j = -\frac{\theta_j}{\left(\theta_0 + \sum_{j=1}^N \theta_j\right) \left(\theta_0 + \sum_{j=1}^{N-1} \theta_j\right)} \quad (19)$$

The threshold of the inverse model is expressed for  $j = 1, 2, \dots, N$  as [3]:

$$s_j = \theta_0 r_j + \sum_{j=1}^{N-1} \theta_j (r_j - r_{j-1}) \quad (20)$$

### 6 Simulation Example

The response characteristics of the generalized Prandtl-Ishlinskii model is evaluated under a harmonic input of the form:  $u(t) = 6 \sin(2\pi t)$ ,  $t \in [0, 2]$ . The simulations are performed using different envelope functions to study their influences on the outputs



**Fig. 2.** Compensation of hysteresis with saturation nonlinearities using the inverse generalized Prandtl-Ishlinskii model as a feedforward compensator. (Square- the inverse generalized Prandtl-Ishlinskii model, Circle- the generalized Prandtl-Ishlinskii model, Triangular- the output of the compensation).

of the generalized play operators. For the illustrative example, hyperbolic-tangent functions are selected for the envelope functions of the generalized play operator, such that:  $f(u) = g(u) = 6 \tanh(0.3u)$ . The results clearly show that the generalized model yields saturated hysteresis loops, which are attributed to the generalized play operator and the selected envelope functions. Equations (13) and (14) of following constants  $\rho = 0.1$ ,  $\tau = 0.1$ , and  $\delta = 0.24$  are selected to obtain the weights and the thresholds of the generalized Prandtl-Ishlinskii model.

Equations (19) and (20) are used to obtain the inverse of the generalized Prandtl-Ishlinskii model to compensate for the hysteresis nonlinearities represented by the generalized Prandtl-Ishlinskii model. As shown in Fig. 2, the results demonstrate linear relation with unity slope, illustrating effective compensation of hysteresis nonlinearities.

## 7 Experimental Verification

The effectiveness of the inverse generalized Prandtl-Ishlinskii model in compensating the hysteresis effects is investigated through laboratory experiments on a piezoceramic actuator. The experiments were performed to compensate the hysteresis loops of the piezoceramic actuator. The generalized Prandtl-Ishlinskii model is applied in order to characterize the hysteresis nonlinearities of the actuator. Fig. 3 illustrates the schematic of the experimental setup including the following elements:

1. **Piezoceramic actuator:** P-753.31C piezoceramic actuator manufactured by Physik Instrumente Company was considered for measurements of the hysteresis properties. The actuator provided a maximum displacement of 100  $\mu\text{m}$  from its static equilibrium position. The excitation voltage to the actuator ranged from 0 to 100 V;
2. **Capacitive sensor:** An integrated capacitive sensor was used for measurements of the actuator displacement response with a sensitivity of 1  $\mu\text{m}/\text{V}$ ;
3. **Voltage amplifier:** The excitation voltage to the actuator was applied through a voltage amplifier (LVPZT, E-505), with a fixed gain of 10;
4. **Data acquisition system:** The actuator displacement response signal was acquired in the dSpace ControlDesk, together with the input signal.

Two different experiments were performed in the laboratory involving characterization of hysteresis and compensation. The initial experiments were performed to characterize the hysteresis properties of the piezoceramic actuator. The measured data were used to identify the generalized model parameters. The inverse generalized model was subsequently identified and applied as a feedforward compensator in the ControlDesk platform. The measurements were then performed to measure the compensated displacement response under the same input. In the second experiment, the input  $u(t) = 45 \sin(\pi t)$  was applied to the actuator through the inverse model. The measured displacement and input voltage data were used for identifying the parameters of the generalized Prandtl-Ishlinskii model subject to the harmonic input used in the

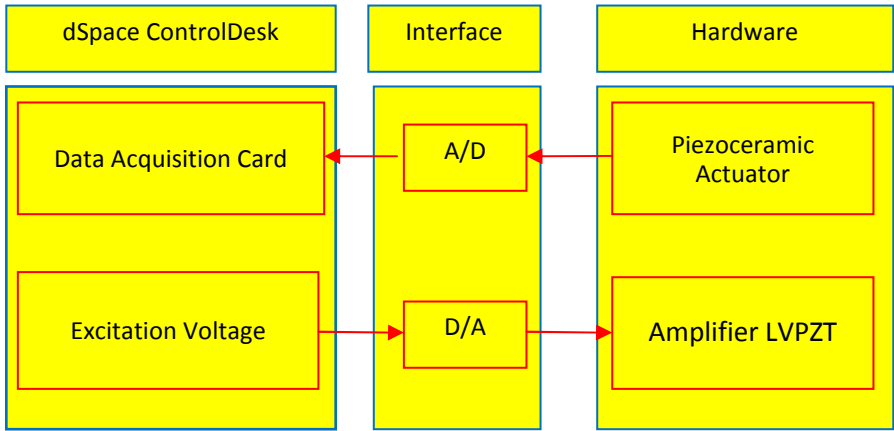


Fig. 3. Experimental setup

experimental study. Considering the nearly symmetric hysteresis properties of the piezoceramic actuator, identical linear envelope functions were chosen, such that:

$$\Gamma_f(t) = \Gamma_g(t) = \xi u(t) \tag{21}$$

where  $\xi$  is a positive constant. The threshold  $r_i$  were chosen as:

$$r_j = \eta j \tag{22}$$

where  $\eta$  is a positive constant. The validity of the generalized Prandtl-Ishlinskii model employing the generalized play operators was examined by comparing the model displacement responses with the measured data.

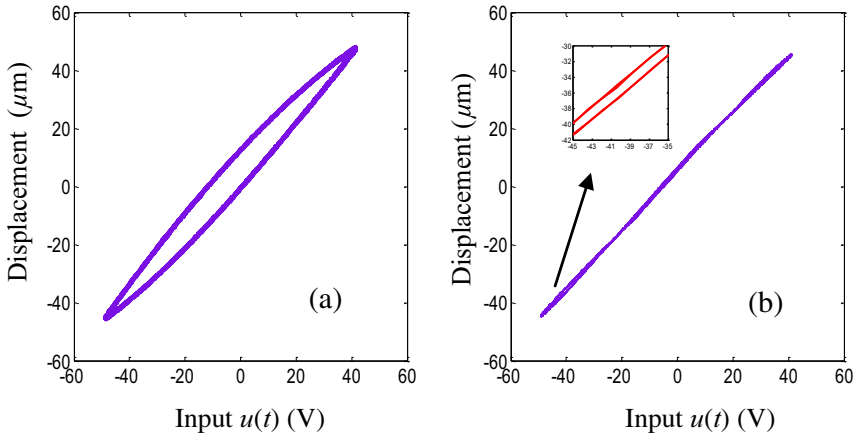


Fig. 4. Measured output displacement of the piezoceramic actuator under the harmonic input voltage: (a) without the inverse generalized Prandtl-Ishlinskii model, and (b) with the inverse generalized Prandtl-Ishlinskii model

The results clearly suggest that the model can effectively predict the hysteresis properties of the piezoceramic actuator. Furthermore, the output-input properties appear to be symmetric and the generalized model can accurately describe the symmetric hysteresis. The inverse of the generalized Prandtl-Ishlinskii hysteresis model was employed as a feedforward controller to compensate for hysteresis nonlinearities of the piezoceramic actuator. The measured output-input characteristics of the piezoceramic actuator with inverse model feedforward compensator are illustrated in Fig. 4(b). The results show that the inverse model can effectively suppress the hysteresis nonlinearities.

## 8 Conclusions

The results show that the integration of the proposed generalized play operator to the Prandtl-Ishlinskii model can effectively characterize the asymmetric hysteresis properties of a class of smart actuators. The results suggest that the feedforward compensator based on inverse generalized Prandtl-Ishlinskii model yields nearly perfect compensation of the effects of hysteresis nonlinearities. The preliminary laboratory experiments conducted to evaluate real-time compensation effectiveness of the inverse generalized model for a piezoceramic actuator revealed peak displacement error in the order of 2%, which was attributed to slight errors in the model.

## References

1. Smith, R.: *Smart Material System: Model Development*. Society for Industrial and Applied Mathematics, Philadelphia (2005)
2. Brokate, M., Sprekels, J.: *Hysteresis and Phase Transitions*. Springer, New York (1996)
3. Krejci, P., Kuhnen, K.: Inverse control of systems with hysteresis and creep. *IEE Proc. Control Theory Application* 148, 185–192 (2001)
4. Al Janaideh, M.: *Generalized Prandtl-Ishlinskii Hysteresis Model and its Analytical Inverse for Compensation of Hysteresis in Smart Actuators*. PhD thesis. Concordia University, Montreal, Canada (2009)
5. Falvo, A., Furgiuele, F., Maletta, C.: Two-way shape memory effect of a Ti rich NiTi alloy: experimental measurements and numerical simulations. *Smart Materials and Structures* 16, 771–778 (2007)

# Hysteresis Modeling of Piezoelectric Actuators Using the Fuzzy System

Pengzhi Li, Guoying Gu, Leijie Lai, and Limin Zhu\*

State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering,  
Shanghai Jiao Tong University, Shanghai 200240, China  
zhulm@sjtu.edu.cn

**Abstract.** An approach of hysteresis modeling in piezoelectric actuators is presented based on the multi-input single-output (MISO) fuzzy system. The proposed model adopts first-order Takagi-Sugeno (T-S) fuzzy system and transforms the multi-valued hysteresis into a one-to-one mapping with the extended input space vector. The generated fuzzy subspaces (multi-dimensional fuzzy sets) assign the maximum membership degree to the input data vectors. Fewer fuzzy subspaces are obtained by introducing the nearest neighbor and super radius concepts. The consequent parameter optimization is implemented after training the fuzzy system. Experimental results demonstrate that this methodology is algorithmically easy and can achieve high modeling accuracy.

**Keywords:** hysteresis, MISO, fuzzy system, parameter optimization.

## 1 Introduction

Piezoelectric actuators are widely applied in high-precision applications [1-3] e.g. micro-robot systems, nanopositioning stages and atomic force microscopes due to their high resolution, large bandwidth, fast response and large output force. However, the inherent hysteresis nonlinearity severely degrades the system performance and results in instability of closed-loop systems. It is a challenging work to model and compensate for the hysteresis.

To characterize the hysteresis, many hysteresis models are developed, which can be mainly classified into two types: physics-based models and phenomenological models. The physics-based approach is inspired from the underlying physics of the actuator or material and is derived based on the empirical observations. Jiles and Thoeke [4] used a set of experimental magnetization measurements and then calculated the hysteresis parameters from the anhysteretic susceptibility at the origin, the coercivity, remanence and the coordinates of the hysteresis loop tip. The phenomenological modeling method employs a mathematical structure to describe hysteresis without considering its fundamental physical characters. Bashash and Jalili [5] proposed a modified Prandtl-Ishlinskii (PI) hysteresis model to improve the modeling accuracy by introducing a new parameter in the primary backlash operators. Janaideh *et al.* [6] proposed a rate-dependent play hysteresis operator and applied it to the classical PI model in

---

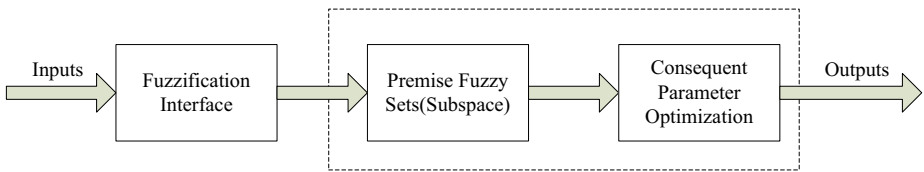
\* Corresponding author.

conjunction with density functions to describe the rate-dependent hysteresis. As a universal function approximator [7], fuzzy systems are widely used to model hysteresis recently. Tafazoli and Demirli [8] adopted the subtractive clustering fuzzy modeling approach to solve the problem of the hysteresis for the airfoil. Azzerboni *et al.* [9] presented an adaptive neuro-fuzzy inference system model of scalar hysteresis to describe symmetric minor loops of a magnetic material. K. Xu *et al.* [10] proposed a method of modeling stress-dependent hysteresis nonlinearity for the giant magnetostrictive actuator based on a modified PI model and fuzzy tree identification.

This paper proposes a simple approach of hysteresis modeling in piezoelectric actuators based on the MISO fuzzy system. Fewer Takagi-Sugeno (T-S) fuzzy rules [11] are generated with simple training algorithm. This method is suitable for modeling a wide range of hysteresis under different excitation signals and can achieve high modeling accuracy. The remainder of this paper is organized as follows. In Section 2, we present the developed fuzzy-based hysteresis model. Then the algorithm of model training and parameters optimization are given in Section 3. In Section 4, we show the experimental validation and simulation results. Section 5 concludes this paper.

## 2 Fuzzy System Model

The basic structure of the proposed fuzzy system consists of two blocks: fuzzification interface and fuzzy inference rules. Fig. 1 shows the block diagram of the fuzzy system. The fuzzification interface defines a mapping from a real-valued space to a fuzzy space and converts a crisp value to a fuzzy number through the definition of membership functions. The block of T-S IF-THEN fuzzy inference rules [11] consists of parts of premise fuzzy sets and consequent parameter optimization. Premise fuzzy sets construct multi-dimensional fuzzy subspaces in IF parts of rules and consequent parameter optimization calculates the coefficients of piecewise linear structure in THEN parts of rules. T-S fuzzy rules have crisp value outputs and thus make defuzzification interface needless.



**Fig. 1.** Structure diagram of fuzzy system

In order to transform the multi-valued hysteresis into a one-to-one mapping, this work uses a MISO fuzzy system with  $m$  inputs and 1 output. Therefore, the corresponding input and output linguistic variables' universes can be written as  $X_i (i=1,2,\dots,m)$  and  $Y$ . In addition, the input space can be defined as

$$X = X_1 \times X_2 \times \dots \times X_m$$

As discussed above, membership function characterizes the fuzzification interface. Triangular, trapezoidal, Gaussian and Sigmoid membership functions are commonly

used in fuzzy system [12]. To construct the proposed model easily, we choose triangular membership function in this work. Subsequently, we define  $c_i$  fuzzy sets  $A_i^1, A_i^2, \dots, A_i^{c_i}$  with triangular membership function in the  $i$ th universe  $X_i$  as follows:

$$\mu_{A_i^c}(x) = \begin{cases} 1 - \frac{|x - a_i^c|}{b_i^c}, & x \in [a_i^c - b_i^c, a_i^c + b_i^c], 1 \leq c \leq c_i, c \in N \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where  $N$ ,  $a_i^c$  and  $b_i^c$  denote the set of all natural numbers, the mean and spread of the fuzzy set respectively.

Suppose the value range of the input variable  $X_i$  in the  $i$ th universe is  $[\alpha_i, \beta_i]$ , we can get

$$b_i^1 = b_i^2 = \dots = b_i^{c_i} = (\beta_i - \alpha_i) / c_i, a_i^c = \alpha_i + (c - 1) \cdot w_i^c, 1 \leq c \leq c_i, c \in N$$

It can be proved that fuzzy sets  $A_i^1, A_i^2, \dots, A_i^{c_i}$  are normal, uniform and complete in  $[\alpha_i, \beta_i]$  [12]. So the fuzzy set of the  $i$ th input variable can be expressed as

$$F_i = \{A_i^1, A_i^2, \dots, A_i^{c_i}\}, 1 \leq i \leq m, i \in N \quad (2)$$

In the model, the number of first-order T-S IF-THEN fuzzy rules used is  $L$ , and the  $l$ th fuzzy rule is expressed as

$$R^l : \text{IF } \mathbf{x} \text{ is } A^l \text{ THEN } y \text{ is } \mathbf{q}_l \cdot \tilde{\mathbf{x}} \quad (3)$$

where  $A^l = [A_1^l, A_2^l, \dots, A_m^l]^T \subset X$ ,  $\mathbf{q}_l = [q_{l1}, q_{l2}, \dots, q_{lm}, q_{l(m+1)}]$ ,  $1 \leq l \leq L, l \in N$ ,

$$\mathbf{x} = [x_1, x_2, \dots, x_m]^T, \tilde{\mathbf{x}} = [\mathbf{x}^T, 1]^T.$$

In addition,  $A_i^l (i = 1, 2, \dots, m)$  is fuzzy set described by equation (1),  $\mathbf{q}_l \cdot \tilde{\mathbf{x}}$  is the consequent crisp real value, and  $q_{l1}, q_{l2}, \dots, q_{lm}, q_{l(m+1)}$  are corresponding crisp real-valued coefficients.

According to equation (1),  $A^l$  can be represented as

$$A^l = \{\mathbf{a}^l, \mathbf{b}^l\} \quad (4)$$

where  $\mathbf{a}^l = [a_1^l, a_2^l, \dots, a_m^l]^T$ ,  $\mathbf{b}^l = [b_1^l, b_2^l, \dots, b_m^l]^T$ . With respect to the input vector

$\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{km}]^T, (k = 1, 2, \dots, n)$ , the membership function  $A^l(\mathbf{x}_k)$  of a multi-dimensional fuzzy set [13] can be defined as follows:

$$A^l(\mathbf{x}_k) = \max \left\{ 1 - \frac{\left[ \sum_{i=1}^m (a_i^l - x_{ki})^2 \right]^{1/2}}{\left[ \sum_{i=1}^m (b_i^l)^2 \right]^{1/2}}, 0 \right\} \quad (5)$$



The  $l$ th fuzzy basis function can be obtained as

$$w^l(\mathbf{x}_k) = \frac{A^l(\mathbf{x}_k)}{\sum_{l=1}^L A^l(\mathbf{x}_k)} \quad (6)$$

Therefore, the weighted average output value of fuzzy system is

$$\hat{y}_k = \sum_{l=1}^L w^l(\mathbf{x}_k) \mathbf{q}_l \cdot \tilde{\mathbf{x}}_k \quad (7)$$

Assuming that all the multi-dimensional fuzzy sets  $A^l$  of the fuzzy system and the corresponding fuzzy basis functions  $w^l(\mathbf{x}_k)$  are known, the optimal  $\mathbf{q}_l$  value can be obtained by minimizing  $E$ .  $E$  is defined in equation (8). We minimize  $E$  via the least squares method based on the input and output sampling data  $(\mathbf{x}_k; y_k)$  ( $k = 1, 2, \dots, n$ ).

$$E = \sum_{k=1}^n \left[ y_k - \sum_{l=1}^L w^l(\mathbf{x}_k) \mathbf{q}_l \cdot \tilde{\mathbf{x}}_k \right]^2 \quad (8)$$

Then the optimal parameter value can be expressed as

$$\mathbf{q} = [W^T W]^{-1} W^T Y_n \quad (9)$$

where

$$\begin{aligned} \mathbf{q} &= [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_L]^T = [q_{11}, q_{12}, \dots, q_{1(m+1)}, q_{21}, q_{22}, \dots, q_{2(m+1)}, \dots, q_{L1}, q_{L2}, \dots, q_{L(m+1)}]^T, \\ W &= [W_1, W_2, \dots, W_n]^T, \quad W_k = [w^1(\mathbf{x}_k) \tilde{\mathbf{x}}_k^T, w^2(\mathbf{x}_k) \tilde{\mathbf{x}}_k^T, \dots, w^L(\mathbf{x}_k) \tilde{\mathbf{x}}_k^T]^T, \\ Y_n &= [y_1, y_2, \dots, y_n]^T, \quad k = 1, 2, \dots, n. \end{aligned}$$

### 3 Fuzzy System Training and Parameter Optimization

Based on the input and output data vector  $(\mathbf{x}_k; y_k)$ , Wang L. X. and Mendal J. M. [14] generated the fuzzy subspace nearest to the input data vector  $\mathbf{x}_k$  using maximum membership method. However, this method obtained too many fuzzy rules which may be mutually conflictive. George T. *et al.* [15] introduced the concept of the nearest neighbor, which effectively reduced the number of generated fuzzy rules by specifying a super radius  $r_0 \in (0.5, 1.0]$ . Due to using zero-order T-S fuzzy rules, the modeling accuracy is a little lower.

Combining concepts of maximum membership and the nearest neighbor aforementioned, we propose an approach of training first-order T-S fuzzy system and optimizing the consequent parameters. The newly proposed approach can significantly reduce the number of fuzzy rules and meanwhile obtain high modeling accuracy. The algorithm can be realized as follows.

Step 1: Use the first input data vector  $\mathbf{x}_1$  to generate its corresponding fuzzy subspace  $A^1(\mathbf{x}_1)$ .  $A^1(\mathbf{x}_1)$  has the same definition as  $A^O(\mathbf{x}_k)$ , which is defined

as  $A^O(\mathbf{x}_k) = [A_1^{O_1}, A_2^{O_2}, \dots, A_m^{O_m}]^T, 1 \leq O \leq L, O \in N, k = 1, 2, \dots, n$ . Besides,  $A_i^{O_i}(x_{ki})$  is subject to the condition of  $\mu_{A_i^{O_i}}(x_{ki}) \geq \mu_{A_i^{O_i'}}(x_{ki})$ , where  $A_i^{O_i} \in F_i, A_i^{O_i'} \in F_i, i = 1, 2, \dots, m$  and  $F_i$  is given in equation (2).

Step 2: Considering subsequent input data vector  $\mathbf{x}_k (k = 2, 3, \dots, n)$ , define

$A^{o'}(\mathbf{x}_k) = \max\{A^l(\mathbf{x}_k)\}$ ,  $l = 1, 2, \dots, L$  where  $L$  represents the number of generated fuzzy subspaces up to present, and  $A^l(\mathbf{x}_k)$  is defined by equation (5). If  $1 - A^{o'}(\mathbf{x}_k)$  is bigger than  $r_0$ , generate a new fuzzy subspace  $A^O(\mathbf{x}_k)$  corresponding to the input data vector  $\mathbf{x}_k$ . Otherwise,  $\mathbf{x}_k$  has a membership degree of some generated fuzzy subspace to a large extent, and do not generate its corresponding new fuzzy subspace any more.

Step 3: After finishing training input data vectors, denote the number of generated fuzzy subspaces as  $L$ , and then calculate  $\mathbf{q}$  using equation (9).

## 4 Experiment

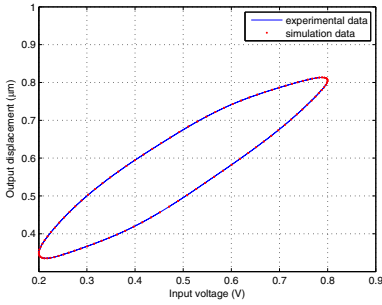
The proposed modeling approach is applied to the experiments on a piezoelectric actuator (PST 150/7/100VS12), which is a preloaded piezo translator (PZT) from Piezomechanik in Germany. The resonance frequency of this actuator is 10 kHz. The actuator provides maximum 90  $\mu m$  displacement and includes an integrated high-resolution strain gauge position sensor (SGS). In order to process the data conveniently, all obtained data are normalized with the maximum values (150 V and 90  $\mu m$  respectively).

The first set of experimental data is collected under the sinusoidal excitation signals. Sampling data include 4-period experimental data, which are normalized in the range of [0, 1]. Two-period data are used to train the fuzzy-based hysteresis model and another two-period data are adopted for model validation. Other modeling parameters are selected as  $c_i = 9, r_0 = 0.505$ , and  $m = 3$ . In addition, suppose that every sampling data pair are  $(x_k, y_k) (k = 1, 2, \dots, n)$ , relative error is given by  $RE = |(y_k - \hat{y}_k) / y_k| \times 100\%$ , where  $\hat{y}_k$  denotes the predictive output based on the proposed model.

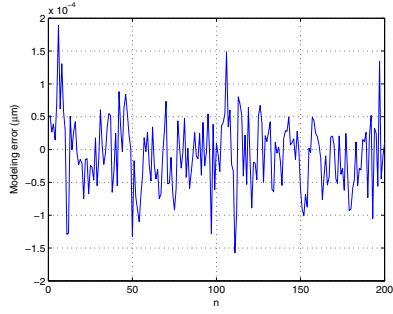
**Table 1.** The modeling maximum relative error and the number of generated fuzzy rules under different sinusoidal excitation signals at different conditions

Conditions	Maximum Relative Error (%)	Number of Fuzzy Rules
$f=100, A=0.3$	0.03	24
$f=50, A=0.3$	0.03	23
$f=1, A=0.3$	0.39	19
$f=100, A=0.4$	0.30	31
$f=100, A=0.2$	1.78	27
$f=100, A=0.05$	0.05	25

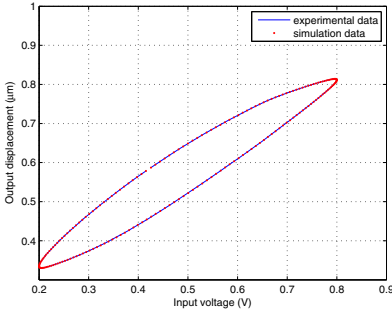
The three-input single-output fuzzy system can be obtained by mapping [16] the data pair to  $((x_{k-1}, y_{k-1}, x_k); y_k)$  ( $k = 2, 3, \dots, n$ ). So the input data vector is assigned as  $\mathbf{x}_k = (x_{k-1}, y_{k-1}, x_k)$ . Fig. 2 and Fig. 3 show the comparisons between the experimental data and simulation results under different sinusoidal excitation signals at different conditions. The modeling maximum relative error and number of fuzzy rules are also illustrated in Table 1. The simulation results show that the T-S fuzzy-based model well describe and predict the rate-dependent hysteresis characterization.



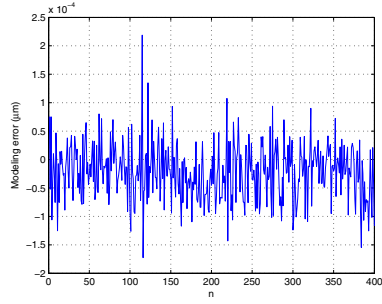
(a) Hysteresis  $f=100$  Hz



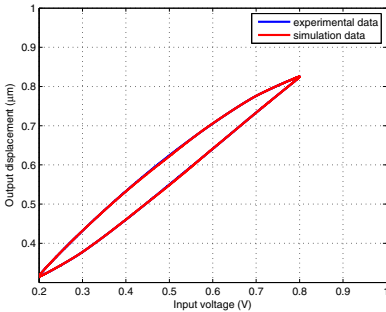
(b) Error  $f=100$  Hz



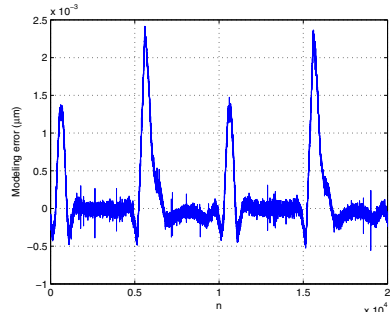
(c) Hysteresis  $f=50$  Hz



(d) Error  $f=50$  Hz

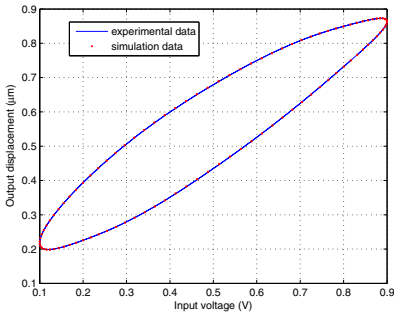


(e) Hysteresis  $f=1$  Hz

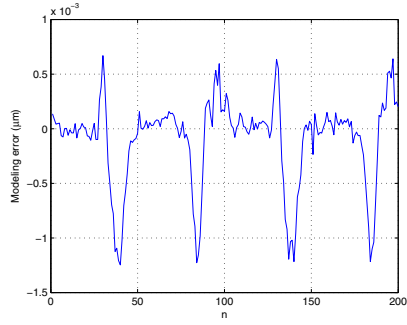


(f) Error  $f=1$  Hz

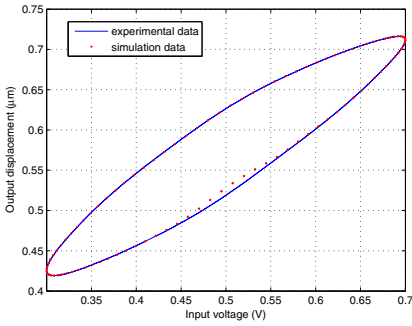
**Fig. 2.** Comparisons between the experimental data and simulation results, and additionally the modeling error under sinusoidal excitation signal  $x(t) = 0.5 + 0.3\sin(2\pi ft)$  at  $f=100, 50, 1$  Hz



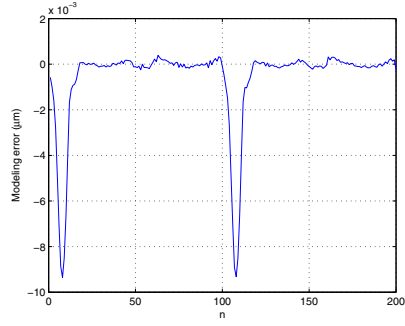
(a) Hysteresis  $A=0.4$



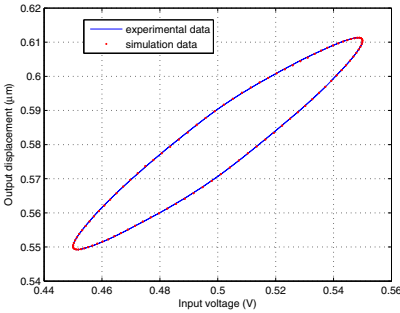
(b) Error  $A=0.4$



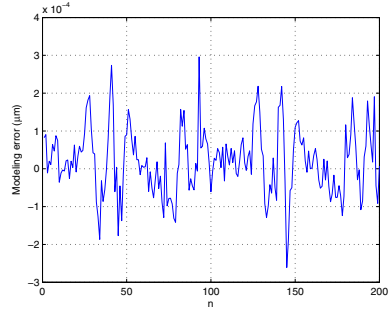
(c) Hysteresis  $A=0.2$



(d) Error  $A=0.2$



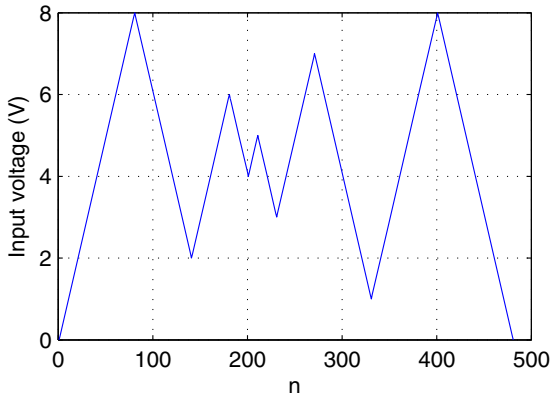
(e) Hysteresis  $A=0.05$



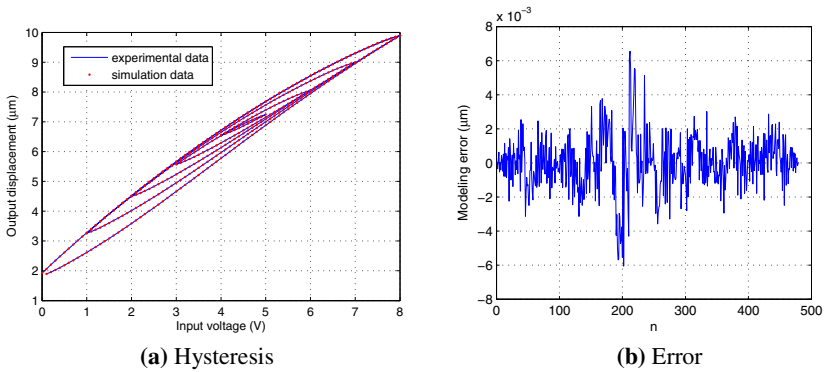
(f) Error  $A=0.05$

**Fig. 3.** Comparisons between the experimental data and simulation results, and additionally the modeling error under sinusoidal excitation signal  $x(t) = 0.5 + A \sin(2\pi 100t)$  at  $A=0.4, 0.2, 0.05$

To compare the modeling performance with the method George T. *et al.* developed, another experiment was done with the non-uniform sawtooth-like excitation signals as shown in Fig.4. In this experiment, the captured data pairs  $(x_k, y_k)$  ( $k=1, 2, \dots, 481$ ) are normalized in the range of  $[0, 10]$ . Fig. 5 (a) shows comparisons between the experimental data and simulation results. In addition, the modeling error is illustrated in Fig. 5 (b). With the same modeling parameters  $c_i=12$ ,



**Fig. 4.** Non-uniform excitation signal



**Fig. 5.** Comparisons between the experimental data and simulation results, and additionally the modeling error

$r_0 = 0.505$ ,  $m=3$ , performance comparisons are given in Table 2. The results show that modeling accuracy proposed by our proposed model is over 20 times higher than George T. *et al.* developed. In other words, to achieve the same modeling accuracy, the proposed model in this paper generates fewer fuzzy rules than George T. *et al.*

In general, different modeling parameters influence the modeling performance within certain boundary. Table 3, Table 4 and Table 5 show the influence of  $c_i$ ,  $r_0$ ,  $m$  respectively. It is obvious that bigger  $c_i$  and smaller  $r_0$  result in higher modeling

**Table 2.** Performance comparisons between models proposed by this paper and George T. *et al.*

Model	Maximum Error	Maximum Relative Error (%)	Number of Fuzzy Rules	Root-mean-square Error
this paper	0.007	0.09	34	2.8e-006
George T.	0.125	3.15	34	1.3e-003

**Table 3.** Modeling performance with respect to  $c_i$ 

$c_i$	Maximum Error	Maximum Relative Error (%)	Number of Fuzzy Rules	Root-mean-square Error
4	0.011	0.40	6	5.8e-006
6	0.011	0.17	12	4.5e-006
8	0.010	0.14	16	3.9e-006
10	0.008	0.11	25	3.1e-006

**Table 4.** Modeling performance with respect to  $r_0$ 

$r_0$	Maximum Error	Maximum Relative Error (%)	Number of Fuzzy Rules	Root-mean-square Error
0.6	0.008	0.11	23	3.2e-006
0.7	0.010	0.13	18	3.8e-006
0.8	0.010	0.15	13	5.1e-006
0.9	0.010	0.14	12	5.3e-006
1.0	0.010	0.14	11	5.4e-006

**Table 5.** Modeling performance with respect to  $m$ 

$m$	Maximum Error	Maximum Relative Error (%)	Number of Fuzzy Rules	Root-mean-square Error
3	0.013	0.38	5	6.7e-006
4	0.174	5.23	5	4.0e-004
5	0.010	0.14	5	3.1e-006
6	0.173	5.21	6	4.0e-004
7	0.010	0.14	7	2.6e-006

accuracy but more fuzzy rules. Meanwhile,  $m$  has a relatively small effect on modeling accuracy and the number of fuzzy rules.

## 5 Conclusion and Future Work

This paper proposes a simple approach of hysteresis modeling based on MISO first-order T-S fuzzy system. The proposed model uses input and output experimental data to train the fuzzy system and optimize the consequent parameters. Experimental results show that the fuzzy-based model can well characterize the hysteresis

nonlinearity with high modeling accuracy. The fuzzy system model has some idiosyncrasy as follows:

- (1). With high modeling accuracy, the algorithm is easily realized and implemented in real-time controllers.
- (2). Only input and output experimental data are adopted in the proposed phenomenological modeling method, which is not involved in the underlying physical hysteresis characteristics. Good expansibility can be attained by changing parameters  $[\alpha_i, \beta_i]$  denoted as the value range of the input data vector.
- (3). Extended into multi-dimensional input space, the multi-valued problem of hysteresis can be solved through sampling data mapping defined by
 
$$\begin{cases} Y_k = y_k \\ \mathbf{x}_k = (x_{k-j}, y_{k-j}, \dots, x_{k-1}, y_{k-1}, x_k) \end{cases}.$$
- (4). Different modeling accuracy can be achieved by changing some parameters such as  $c_i, r_0$  and  $m$ .

The future work will be focused on the real-time controller design for tracking control of the piezoelectric actuator using the developed model. We will also have some try on fuzzy model-based predictive control.

## Acknowledgments

This work was partially supported by the Science and Technology Commission of Shanghai Municipality under Grant Nos. 09520701700 and 09JC1408300, and the National Key Basic Research Program under Grant No. 2007CB714005.

## Reference

1. Ang, W.T., Khosla, P.K., Riviere, C.N.: Feedforward controller with inverse rate-dependent model for piezoelectric actuators in trajectory-tracking applications. *IEEE/ASME Transactions on Mechatronics* 12(2), 134–142 (2007)
2. Aphale, S.S., Devasia, S., Moheimani, S.O.R.: High-bandwidth control of a piezoelectric nanopositioning stage in the presence of plant uncertainties. *Nanotechnology* 19(12), 125503 (2008)
3. Hu, H., Georgiu, H.M.S., Benmrad, R.: Enhancement of tracking ability in piezoceramic actuators subject to dynamic excitation conditions. *IEEE/ASME Transactions on Mechatronics* 10(2), 230–239 (2005)
4. Jiles, D.C., Thoelke, J.B.: Theory of ferromagnetic hysteresis determination of model parameters from experimental hysteresis loops. *IEEE Transactions on Magnetics* 25(5), 3928–3930 (1989)
5. Bashash, S., Jalili, N.: Robust multiple frequency trajectory tracking control of piezoelectrically driven micro nanopositioning systems. *IEEE Transactions on Control Systems Technology* 15, 867–878 (2007)
6. Al Janaideh, M., Rakheja, S., Su, C.Y.: Experimental characterization and modeling of rate-dependent hysteresis of a piezoceramic actuator. *Mechatronics* 19(5), 656–670 (2009)

7. Wang, L.X., Mendel, J.M.: Fuzzy basis functions, universal approximation, and orthogonal least squares learning. *IEEE Transactions on Neural Networks* 3(5), 807–814 (1992)
8. Tafazoli, M., Demirli, K.: Fuzzy modeling of hysteresis from input-output data. In: *IFSA World Congress and 20th NAFIPS International Conference*, pp. 3009–3014 (2001)
9. Azzerboni, B., Carpentieri, M., Finocchio, G., La Foresta, F.: A fuzzy model of scalar hysteresis on soft magnetic materials. *Physica B: Condensed Matter* 343(1-4), 132–136 (2004)
10. Xu, K., Zhang, Z., Mao, J.: Modeling of stress dependent hysteresis nonlinearity based on fuzzy tree for GMA 2008. In: *Proceedings of the IEEE International Conference on Automation and Logistics*, pp. 331–335 (2008)
11. Sugeno, M., Yasukawa, T.: A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems* 1(1), 7–31 (1993)
12. Wang, L.X.: *Adaptive fuzzy systems and control*. Prentice-Hall, Englewood Cliffs (1994)
13. Nie, J.: Fuzzy control of multivariable nonlinear servomechanisms with explicit decoupling scheme. *IEEE Transactions on Fuzzy Systems* 5(2), 304–311 (1997)
14. Wang, L.X., Mendel, J.M.: Generating fuzzy rules from numerical data with applications. *IEEE Transactions on System, Man and Cybernetics* 22(6), 1414–1427 (1992)
15. George, T., Haralambos, S., George, B.: A simple algorithm for training fuzzy systems using input–output data. *Advances in Engineering Software* 34, 247–259 (2003)
16. An, J., Yang, Q., Ma, Z.: Study on method of modeling and controlling of magnetostrictive material. In: *17th International Zurich Symposium on Electromagnetic Compatibility*, pp. 387–390. IEEE, Piscataway (2006)



# On PSO Based Bouc-Wen Modeling for Piezoelectric Actuator

ZhenYan Wang<sup>1,2</sup> and JianQin Mao<sup>1</sup>

<sup>1</sup> School of Automation Science and Electrical Engineering,  
Beijing University of Aeronautics and Astronautics,  
100191, Beijing, P.R. China

<sup>2</sup> School of Electronical and Information Engineering,  
Taiyuan University of Science and Technology,  
030024, Taiyuan, P.R. China  
w9851@126.com, jqmao@buaa.edu.cn

**Abstract.** The paper proposes an effective modeling method for piezoelectric actuator with hysteresis, in which Bouc-Wen model is employed to describe the piezoelectric hysteretic phenomenon. In order to determine the model parameters, the particle swarm optimization (PSO) algorithm is utilized. Finally, the comparisons of numerical simulations and experimental results for the piezoelectric actuator in different frequency are drawn. It is found that the Bouc-Wen model based on intelligent optimization can model excellently the hysteretic phenomenon to achieve high-quality fitting curve between experimental and model data. The model generation performance and hysteresis curve fitting ability at 20-80 Hz prove the effectiveness of this method.

**Keywords:** piezoelectric actuator; hysteresis; PSO; Bouc-Wen model.

## 1 Introduction

The piezoelectric actuator (PZT) is widely used to meet some specific requirements, especially in the field of precision pointing[1], micro-manipulation[2], micro-robot arm[3], and vibration active control[4] for the prominent advantages such as large force generation, high stiffness, high control precision, low power consumption and fast response. Nevertheless, it is well known that the main drawback of PZT comes from the nonlinearities mainly attributed to hysteresis behavior, creep phenomena and high frequency vibration. As the uppermost nonlinear characteristics, hysteresis behavior restricts heavily its application in some high-precision required task and situation.

Hysteresis is universal nonlinear effect in smart materials, such as piezoelectrics, magnetostrictives, shape memory alloys, and electroactive polymers which is commonly viewed as an undesirable, detrimental effect in engineering systems [5]. The piezoelectric actuator exhibits hysteresis phenomena between the applied voltage and output displacement, which makes a challenge to control its movements. Therefore, it is crucial to model precisely for PZT with taking into account the hysteresis to facilitate its application in micro-manipulation field.

In order to solve the hysteresis nonlinear problem in high-precision applications, many modeling methods for the hysteresis behavior have been researched widely [6-10]. Typically, the hysteresis phenomenon is described by the nonlinear model such as the Preisach model [6], neural network model [7], NARMAX model [8], Prandtl-Ishlinskii model [9], and Bouc-Wen model [10-11], etc. By the weighted superposition of delayed relays, the Preisach model can model complex hysteresis loops with a large number of model parameters to produce a good fit with experimental data [6]. With a special hysteretic operator, the multi-valued mapping of hysteresis is transformed into a one-to-one mapping, so that neural networks can be applied to the approximation of the hysteresis behavior in [7]. The Bouc-Wen model approach is popular in actual application [12-13] because of its simple interpretation as a nonlinear black-box model and the relatively low number of parameters needed. So in this paper the Bouc-Wen model is utilized to approximate the hysteretic nonlinear of the piezoelectric actuator. As the key part to implement hysteretic nonlinear fitting, Bouc-Wen model parameters identification has been widely discussed and researched [14-17]. Both adaptive estimation approach [14-15] and evolutionary algorithm [13, 16-17] had been investigated to identify model parameters from experimental data.

Particle swarm optimization (PSO) algorithm, developed by Kennedy and Eberhart in 1995 [18], is an evolutionary algorithm which is inspired by the mechanism of biological swarm social behavior such as fish schooling and bird flocking. It differs from other evolutionary techniques in the adoption of velocity of individuals, and it can search randomly more than genetic algorithm (GA) and avoid falling into the local optimum with faster convergence speed. Due to its simple principles and low computational complexity, PSO algorithm has been widely applied in a wide domain in function optimizations such areas as optimal circuit design and data clustering [19-20]. In this paper, in order to model the hysteretic curve between the displacement output and applied voltage of a type of piezoelectric micro-actuators, the Bouc-Wen model is adopted to characterize the nonlinear hysteretic relationship. The corresponding parameter identification based on PSO algorithm is developed. Experimental setup is established to evaluate the effectiveness of the proposed method.

The organization of this paper is as follows. In Section 2, The PZT experimental setup for data acquisition between input voltage and output displacement is explained. The Bouc-Wen hysteresis model for the piezoelectric actuator is introduced in Section 3. The modeling method with PSO parameter identification is developed in Section 4. In section 5, comparisons between numerical simulations and experimental results are shown. Finally, some conclusions and prospective trend of hysteresis modeling for PZT with Bouc-Wen are drawn.

## 2 Experimental Setup

For a PZT, the relation between the displacement and applied voltage is not linear but is a nonlinear hysteretic characteristic. And the hysteretic loop with different frequency of PZT will be measured and described by an experimental setup. The schematic diagram of experiment for data acquisition is shown in Fig.1, while the photograph of the whole experimental setup is shown in Fig.2. The whole experimental system includes a piezoelectric actuator produced by PIEZOMECHANIK GmbH,

a piezoelectric ceramics drive power, an eddy current sensor(85745) and the DSPACE(DS1103) controller installed in the industrial control computer(IPC) bus slots which is used to realize the real-time experimental input and output data acquisition with a sampling frequency 1000hz.

The applied voltage to actuator is produced by a signal generation module of Simulink. The D/A converter is used to transform the waveform to the specified power amplifier, which has the voltage output range 0~150V. An eddy current sensor is utilized to measure the actual displacement of the PZT, and the corresponding transfer ratio of displacement to the voltage measured from sensor is  $8\text{mV}/\mu\text{m}$ . Finally, the voltage signal of displacement is transformed via the A/D converter to DSPACE control card and recorded in the IPC. The measured hysteresis loops with different frequency are also described in Fig.3. Obviously, the voltage/displacement curves are all hysteresis loops; therefore, it is necessary to investigate further the nonlinear actuator model.

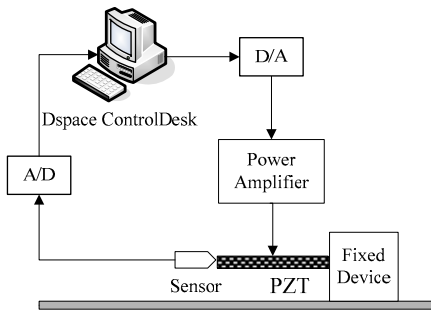


Fig. 1. The schematic diagram

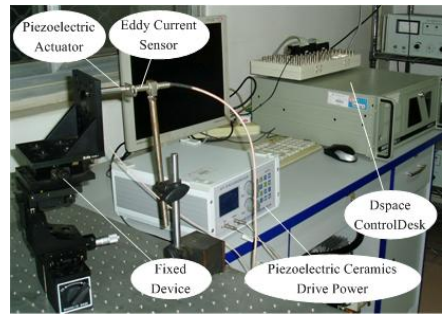


Fig. 2. The photograph of experimental setup

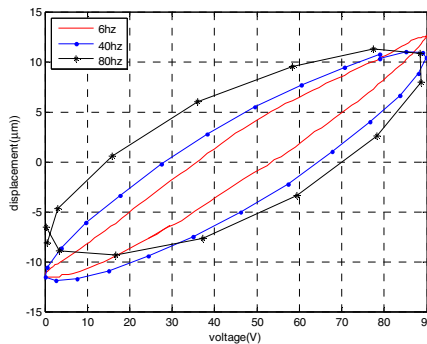


Fig. 3. The hysteresis loop with different frequency

### 3 Bouc-Wen Hysteresis Model

In this paper, the Bouc-Wen model is employed to describe the nonlinear hysteresis. The dynamic model of hysteresis should be established by the first-order nonlinear differential equations [10-11]:

$$\dot{h} = \alpha \dot{V} - \beta |\dot{V}| |h| |h|^{n-1} - \gamma \dot{V} |h|^n. \quad (1)$$

where  $V$  is the applied voltage and  $h$  represents the hysteretic nonlinear term;  $\dot{V}$  and  $\dot{h}$  are the derivatives of  $V$  and  $h$  with respect to time  $t$ , respectively; The coefficient  $\alpha$  controls the amplitude of hysteretic loop, while the coefficients  $\beta$  and  $\gamma$  control the shape of the hysteresis loop. Additionally, the coefficient  $n$  controls the smoothness of transition from elastic to plastic responses. Through appropriate choices of parameters in the model, it can show a wide variety of hysteresis loops with various shapes.

The relation between the piezoelectric actuator displacement and the input voltage is not a linear behavior since the hysteresis effects is considered. When the voltage is applied to the plant, the piezoelectric actuator displacement  $x$  will form the hysteresis loop with respect to the applied voltage as shown in Fig.3. The Bouc–Wen model is considered in this work to model the piezoelectric actuator since it has fewer parameters and is easier to implement. It has already been verified that the Bouc–Wen model is suitable to describe the hysteresis phenomenon of the PZT by Low and Guo [12]. The entire dynamic model between the displacement and the input voltage of the micro-piezoelectric actuator with nonlinear hysteresis is established with Bouc–Wen model as follows:

$$\begin{cases} x = dV - h \\ \dot{h} = \alpha \dot{V} - \beta |\dot{V}| |h| |h|^{n-1} - \gamma \dot{V} |h|^n \end{cases} \quad (2)$$

Where the displacement  $x$  of PZT is represented with a linear term  $dV$  and a nonlinear hysteretic term  $h$ .  $d$  is the ratio of the displacement vs. the applied voltage. This model reflects local history dependence through introducing the nonlinear term  $h$ .

In this paper, the nonlinear model (2) will approximate the actual hysteretic loop with appropriate parameters. The intelligent evolution PSO algorithm is introduced to optimize model parameters which will be introduced in the next section.

## 4 System Modeling Based on PSO Algorithm

### 4.1 PSO Algorithm

The particle swarm optimization technique is one of the most powerful methods for solving unconstrained and constrained global optimization problems[18]. It consists of a number of individuals that denote particles to simulate social behavior ‘flying’ around in a multidimensional search space. The individuals thus have a position and a velocity. The particles evaluate and update their positions with a fitness value at each iteration. By attracting the particles to a better positions with good solutions, each particle remembers its own previously best found position, and particles in the group share memories of their “best” positions, then use those memories to adjust their own velocities, and thus subsequent positions.

Similar to other evolutionary algorithm, PSO algorithm works well by the fitness value of each particle based on the notations of “group” and “evolution”. In the

original PSO, the position of each particle in the swarm represents a possible solution to a problem in D-dimensional space. The position of particle  $i$  ( $i=1\sim N$ ,  $N$  denotes group scale) represents as  $x_i = (x_{i1}, x_{i2}, \dots, x_{id}, \dots, x_{iD})$ . Each particle also maintains a memory of its previous best position  $P_i = (p_{i1}, p_{i2}, \dots, p_{id}, \dots, p_{iD})$ , and a flying velocity along each dimension, represented as  $v_i = (v_{i1}, v_{i2}, \dots, v_{id}, \dots, v_{iD})$ . At each iteration, the distance between the position of the best particle in the swarm  $p_{gd}$  and the current particle position  $x_{id}$ , and the distance between the particle previous best position  $p_{id}$  and the current particle position  $x_{id}$ , are combined to adjust the velocity along each dimension for the particle [18]:

$$v_{id}(t+1) = w v_{id}(t) + c_1 r_1 (p_{id} - x_{id}(t)) + c_2 r_2 (p_{gd} - x_{id}(t)). \tag{3}$$

where  $w$  is the inertia weight and  $t$  means the current iteration number. Acceleration constants  $c_1$  and  $c_2$  are cognitive and social parameters respectively, and  $r_1$  and  $r_2$  are two random values in the range of  $[0, 1]$ . The position  $x_{id}$  is thus iteratively updated with the above velocity as follows [18]:

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \tag{4}$$

The above deterministic and probabilistic parameters reflect the effects on the particle positions from both the individual memory and swarm influence. The portion of the adjustment to the velocity influenced by the previous best position is considered the cognition component, and the portion influenced by the best in the swarm is the social component. Additionally, the position  $x_{id}$  and velocity  $v_{id}$  are usually refined with limits  $x_{id} \in [x_{\min}, x_{\max}]$ ,  $v_{id} \in [v_{\min}, v_{\max}]$  so as to reduce the possibility of the particle moving away from the search space. From (3) and (4), it can be seen that taking advantage of collective intelligence is the distinguishing property of the PSO method.

The inertia weight operator  $w$  plays the role of balancing the global search and the local search; suitable selection of  $w$  can improve the convergence performance of PSO algorithm with less iteration on average to find a sufficiently optimal solution. In order to assure the initial global search and the later local research, a time-varying inertia weight  $w$  is employed[17]:

$$w = w_{\max} - \frac{t}{T}(w_{\max} - w_{\min}). \tag{5}$$

Where  $T$  is the maximum iteration generation.  $w$  decreases linearly from maximum value  $w_{\max}$  to minimum value  $w_{\min}$  during the optimization progress.

### 4.2 Model Identification Based on PSO

Based on PSO algorithm, the Bouc-Wen model parameters  $\alpha, \beta, \gamma, n$  and piezoelectric constant  $d$  are to be identified simultaneously. In application of the PSO method, how to define the fitness function is the key point.

The cost function between experimental data and model described by the root-mean-square error (RMSE) or relative error (RE) are shown in equation (6).

$$J_{RMSE} = \sqrt{\sum_{i=1}^L (Y_{\text{exp}}^i - Y_{BW}^i)^2 / L} \quad , \quad J_{RE} = \sqrt{\sum_{i=1}^L (Y_{\text{exp}}^i - Y_{BW}^i)^2 / \sum_{i=1}^L (Y_{\text{exp}}^i)^2} \quad (6)$$

Where  $Y_{\text{exp}}^i$  is the measured data from experiment at the  $i$ th sampling time,  $Y_{BW}^i$  is the corresponding Bouc-Wen model estimator output, and  $L$  is the total number of samples. In order to fit the model output and experimental data accurately, the fitness functions

$$f(\alpha, \beta, \gamma, n, d) = 1 / (J_{RE} + 1) \quad (7)$$

is utilized as evaluating indicator to verify the model performance.

Since experimental input and output measurements are taken at discrete time intervals  $\Delta t$ , a discrete version model corresponding to (2) of the system is deduced. Therefore, the problem can be described as follows:

$$\begin{aligned} \min \quad & f(\alpha, \beta, \gamma, n, d) \\ \text{subject to} \quad & \begin{cases} y_k = dx_k - h_k \\ h_k = h_{k-1} + \Delta t (\alpha \dot{x}_k - \beta |\dot{x}_k| |h_{k-1}| |h_{k-1}|^{(n-1)} - \gamma \dot{x}_k |h_{k-1}|^n) \end{cases} \end{aligned} \quad (8)$$

The discrete-time form of Bouc-Wen model with five unknown model parameters  $\alpha, \beta, \gamma, n, d$  is employed to model the piezoelectric actuator.

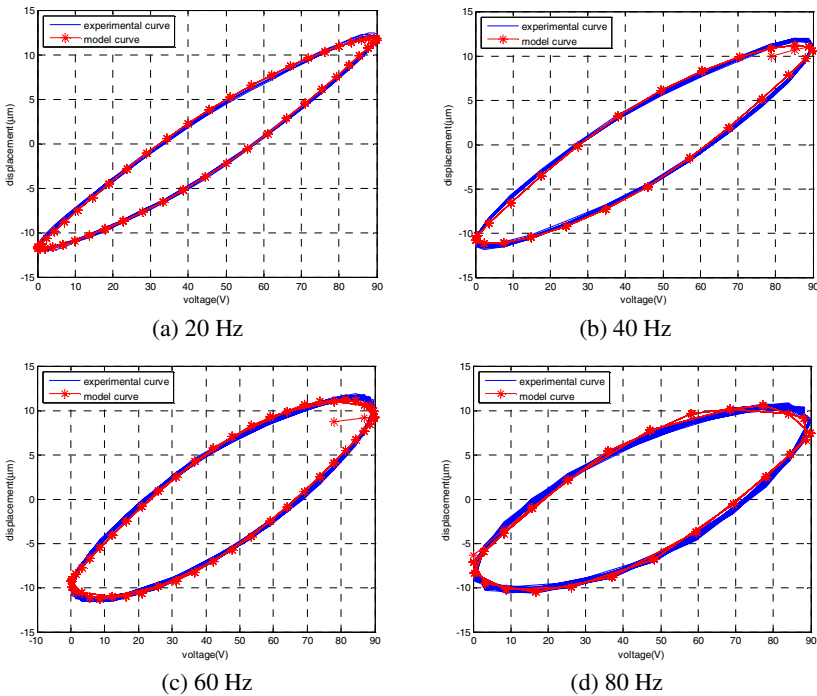
### 4.3 Modeling Steps

- (1) *Data collection and preprocessing*: Measure and record the experimental data and preprocess those data to get the input-output data pairs as training set  $\{V_1, X_1\}$  and testing set  $\{V_2, X_2\}$  respectively.
- (2) *Initialization*: Initialize a group of random particles including model parameters  $\alpha, \beta, \gamma, n, d$  and the PSO parameters such as  $c_1, c_2, w_{\max}$  and  $w_{\min}$ , group size, random position and velocity, and initial vector.
- (3) *Calculate model output and evaluate the fitness value*: Calculate the model output with the initialized parameter  $\alpha, \beta, \gamma, n, d$  according to (8); as well as evaluate the fitness value of each particle according to fitness function (7).
- (4) *Update the best position of individual and global position*: For each particle, by comparing the individual fitness value at present with the individual best position  $p_{\text{best}}$  and the group best position  $g_{\text{best}}$  in the past, the individual and global best position,  $p_{\text{best}}$  and  $g_{\text{best}}$ , are updated if the present value is better than the past.
- (5) *update velocity and position of particles*: Calculate velocity and position of particles according to (3-4).
- (6) *Termination*: Return (3) if termination condition that the best searching positions so far reach the predetermined minimum fitness is not arrived or the maximum iterative number is not reached.

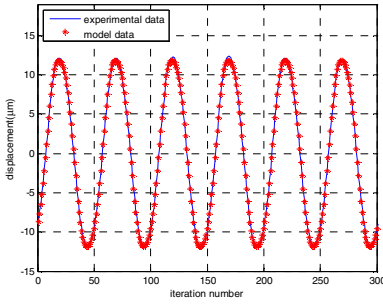
## 5 Numerical Simulation

In this section, modeling results are compared with the experimental data to reveal the performance of the Bouc-Wen hysteresis model. The model identification is performed using the sine wave with the frequency of 20Hz, 40Hz, 60Hz and 80Hz. Two sets of input and output signal of PZT with respect to frequency are sampled as the data pairs. For the parameter identification, in PSO we have taken the acceleration constants  $c_1 = c_2 = 2$ , the group scale  $N=30$  and inertia weighting factor restricts  $w \in [0, 1.2]$  for obtaining best results.

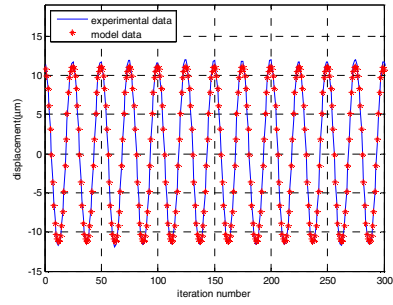
Fig.4 indicates comparison between hysteretic loop measured by experiments and those simulated based on the Bouc-Wen hysteretic model with (8) at 20Hz, 40Hz, 60Hz and 80Hz respectively. Fig.5 is the corresponding model output displacement comparisons at each frequency. We can see that the Bouc-Wen model can reach distinctly the actual hysteretic output displacement to achieve beautiful fitting results. Table 1 is the identified optimal parameters with Bouc-Wen model, and the mean square error and relative error of the training set  $\{V_1, X_1\}$  and testing set  $\{V_2, X_2\}$  defined by (6) are shown in Table 2 to validate the generation ability of Bouc-Wen hysteresis model (2).



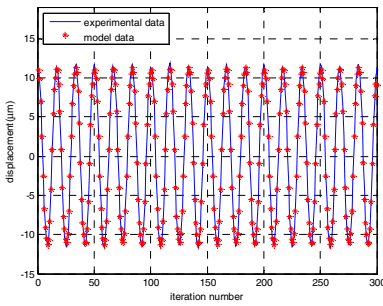
**Fig. 4.** Voltage/displacement hysteresis curve comparison between experimental measurement and the Bouc-Wen model with applied voltage at (a)20Hz, (b)40Hz, (c)60Hz and (d)80Hz, respectively



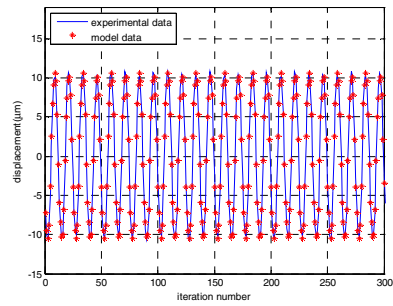
(a) 20 Hz



(b) 40 Hz



(c) 60 Hz



(d) 80 Hz

**Fig. 5.** Iteration/displacement number comparison between experimental measurement and the Bouc-Wen model with applied voltage at (a)20Hz, (b)40Hz, (c)60Hz and (d)80Hz, respectively

**Table 1.** The optimal parameters of Bouc-Wen model

$f$ (HZ)	20	40	60	80
$\alpha$	3.8404e-005	4.1874e-005	4.6330e-005	4.6918e-005
$\beta$	9.9795	9.9969	8.7802	10.4524
$\gamma$	1.8997	2.9930	1.3800	3.5992
$n$	1.1427	1.1339	1.1285	1.1409
$d$ (m/V)	7.2801e-005	7.0718e-005	7.3347e-005	6.5769e-005

**Table 2.** The generation performance of Bouc-Wen model with the training and testing data pairs

$f$ (HZ)	20	40	60	80
$RMSE(\mu m)$ of training	0.2845	0.5544	0.5574	0.8100
$RE$ of training	0.0328	0.0659	0.0670	0.1058
$RMSE(\mu m)$ of testing	0.2530	0.4562	0.5248	0.6974
$RE$ of testing	0.0291	0.0544	0.0631	0.0911



## 6 Conclusions

This study proposed a modeling method for the piezoelectric actuator with hysteresis. By identification research, it is possible to construct a Bouc-Wen model to describe the piezoelectric actuator hysteresis with respect to single-frequency input voltage. The predict ability on the testing data successfully demonstrates that the Bouc-Wen model based PSO algorithm can be utilized to describe the piezoelectric actuator. According to the comparisons between the identified results and the experimental results prove the effectiveness of this method.

The main advantage of Bouc-Wen model is in its simplicity with little parameters and its ability to describe a variety of complicated hysteresis loop. However, it is worth to notice that the proposed modeling scheme is available for the single frequency only at present; so it is necessary to extend or modify the Bouc-Wen model with evolutionary algorithm to model piezoelectric actuators when the rate-dependent characteristic is considered. On the other hand, considering the restriction that the Bouc-Wen model cannot model asymmetry hysteretic loop, the modeling error would not be ignored when specific precision pointing or micromanipulation are needed. These would be the prospective work need to be considered in the future.

## Acknowledgement

This work was supported by the state Key-Program of National Natural Science Foundation of P.R.China (“Research on Control Principles of Operation Stabilization and Maneuvering Ability for Hypersonic Vehicles Based on a Class of Smart Structure”) and the Specialized Scientific Research Doctor Fund of Colleges and Universities (20070006060).

## References

1. Lin, C.J., Yang, S.R.: Precise Positioning of Piezo-Actuated Stages using Hysteresis-Observer Based Control. *Mechatronics* 16, 417–426 (2006)
2. Goldfarb, M., Celanovic, N.: Modeling Piezoelectric Stack Actuators for Control of Micromanipulation. *IEEE Control Systems Magazine* 17(3), 69–79 (1997)
3. Simu, U., Johansson, S.: Evaluation of a Monolithic Piezoelectric Drive Unit for a Miniature Robot. *Sensors and Actuators A: Physical* 101, 175–184 (2002)
4. Viswamurthy, S.R., Ganguli, R.: Modeling and Compensation of Piezoceramic Actuator Hysteresis for Helicopter Vibration Control. *Sensors and Actuators A: Phys.* 135(2), 801–810 (2007)
5. Tan, X., Iyer, R.V.: Modeling and Control of Hysteresis. *IEEE Control Systems Magazine* 29(1), 26–29 (2009)
6. Mayergoyz, I.: *Mathematical Models of Hysteresis*. Springer, New York (1991)
7. Zhang, X.L., Tan, Y.H., Su, M.Y.: Modeling of hysteresis in piezoelectric actuators using neural networks. *Mechanical Systems and Signal Processing* 23(8), 2699–2711 (2009)
8. Deng, L., Tan, Y.H.: Modeling Hysteresis in Piezoelectric Actuators using NARMAX Models. *Sensors and Actuators A: Physical* 149, 106–112 (2009)

9. Kuhnen, K.: Modeling, Identification and Compensation of Complex Hysteresis Nonlinearity: A Modified Prandtl-Ishlinskii Approach. *European Journal of Control* 9(4), 407–418 (2003)
10. Bouc, R.: Forced Vibration of Mechanical Systems with Hysteresis. In: *Proc. of the 4th Conference on Nonlinear Oscillations, Prague, Czechoslovakia (1967)*
11. Wen, Y.K.: Method for Random Vibration of Hysteresis Systems. *Journal of Engineering Mechanics Division* 102(2), 249–263 (1976)
12. Low, T.S., Guo, W.: Modeling of a Three-Layer Piezoelectric Bimorph Beam with Hysteresis. *Journal of Microelectromechanical Systems* 4(4), 230–237 (1995)
13. Lin, C.J., Chen, S.Y.: Evolutionary Algorithm based Feedforward Control for Contouring of a Biaxial Piezo-actuated Stage. *Mechatronics* 19, 829–839 (2009)
14. Chassiakos, A.G., Masri, S.F., Smyth, A.W., Caughey, T.K.: On-Line Identification of Hysteretic Systems. *Journal of Applied Mechanics* 65, 194–203 (1998)
15. Ha, J.L., Fung, R.F., Yang, C.H.: Hysteresis Identification and Dynamic Responses of the Impact Drive Mechanism. *Journal of Sound and Vibration* 283, 943–956 (2005)
16. Charalampakis, A.E., Koumoussis, V.K.: Identification of Bouc–Wen Hysteretic Systems by a Hybrid Evolutionary Algorithm. *Journal of Sound and Vibration* 314, 571–585 (2008)
17. Kao, C.C., Fung, R.F.: Using the Modified PSO Method to Identify a Scott-Russell Mechanism Actuated by a Piezoelectric Element. *Mechanical Systems and Signal Processing* 23, 1652–1661 (2009)
18. Kennedy, J., Eberhart, R.C.: Particle Swarm Optimization. In: *Proceedings of the IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948. IEEE Press, Los Alamitos (1995)*
19. Tawdross, P., Konig, A.: Investigation of Particle Swarm Optimization for Dynamic Reconfiguration of Field-programmable Analog Circuits. In: *Proceedings of the International Conference on Hybrid Intelligent Systems, Rio de Janeiro, Brazil, pp. 259–264 (2005)*
20. Merwe, V.D., Engelbrecht, A.P.: Data Clustering using Particle Swarm Optimization. In: *Proceedings of the IEEE Congress on Evolutionary Computation, Canberra, Australia, pp. 215–220 (2003)*

# Operator-Based Robust Nonlinear Control for Ionic Polymer Metal Composite with Uncertainties and Hysteresis

Aihui Wang, Mingcong Deng, and Dongyun Wang

School of Electronic Information, Zhongyuan University of Technology  
41 Zhongyuan Road, Zhengzhou, 450007, China  
Graduate School of Natural Science and Technology, Okayama University  
3-1-1 Tsushima-Naka, Okayama 700-8530, Japan  
deng@suri.sys.okayama-u.ac.jp

**Abstract.** The ionic polymer metal composite (IPMC) belongs to the category electroactive polymers (EAP), many potential applications for low-mass high-displacement actuators in biomedical and robotic systems have been shown. But identification of some physical parameters for nonlinear IPMC models is still a difficult issue. Moreover, hysteretic behavior exists in IPMCs and affects the performance of actuators, even makes the system with these actuators exhibit undesirable oscillations and instability. In this paper, a new nonlinear model of the IPMC with uncertainties and hysteresis is obtained. According to hysteresis and uncertainties for the proposed model, a nonlinear robust control using operator-based robust right coprime factorization is designed for the IPMC. The effectiveness of the proposed method is confirmed through simulation and experiment.

**Keywords:** IPMC, Uncertainties, Hysteresis, Nonlinear Control, Right Coprime Factorization, Robust Stability.

## 1 Introduction

Various electroactive polymers (EAP) materials, also called artificial muscles, are being developed to enable effective, miniature, light and low power actuators. The Ionic Polymer Metal Composite (IPMC) belongs to the category electroactive polymers (EAP). An IPMC sample consists of a thin ion-exchange membrane (e.g., Nafion) plated on both surfaces with a noble metal as electrodes. Because IPMCs are capable of producing large deformation under a low driving voltage, they have been shown to have many potential applications as biomimetic robotic distributed sensors, actuators, transducers, artificial muscles and so on [1, 2].

Several linear and nonlinear models have been proposed in precision displacement control [3, 4] The linear models are often obtained from approximate method, for linear models, linear quadratic regulator (LQR), proportional integral and derivative (PID), adaptive fuzzy algorithm and impedance control

scheme are usually used in position control for linear models [5, 6]. The IPMC mainly shows nonlinear behaviors, but identification of some physical parameters is difficult in practice. Some physical parameters are small enough the influence for displacement deformation in practice, which can be ignored and considered as uncertainties in the model. According to variables ignored and measurement error of parameters, an improved nonlinear model with uncertainties is obtained. Moreover, hysteretic behavior exists in IPMCs and affects the performance of actuators [7], even makes the system with these actuators exhibit undesirable oscillations and instability. In order to make hysteresis model closer to the real hysteretic behavior, a symmetric PI hysteresis model which is described by symmetric play hysteresis operator with unknown slopes is given.

In precision position control, IPMC actuator has to move from one specified position to another, and has to maintain the position constant. It needs a skilful operator to control manually based on his or her experiences to stop the swing immediately at the right position. To resolve this problem, operator-based robust right coprime factorization is proposed in this paper. It is well known that coprime factorization has been a promising approach for analysis, design, stabilization and control of nonlinear system [8, 9]. Especially, robust right coprime factorization has attracted much attention due to its convenient in researching input-output stability problems of nonlinear system with uncertainties [10, 11, 12, 13]. As a result, there exists hysteretic behavior and some uncertainties for the proposed nonlinear model, the nonlinear IPMC position control system is considered by using robust right coprime factorization approach in this paper, robust right coprime factorization and its application to a nonlinear IPMC control setup are investigated.

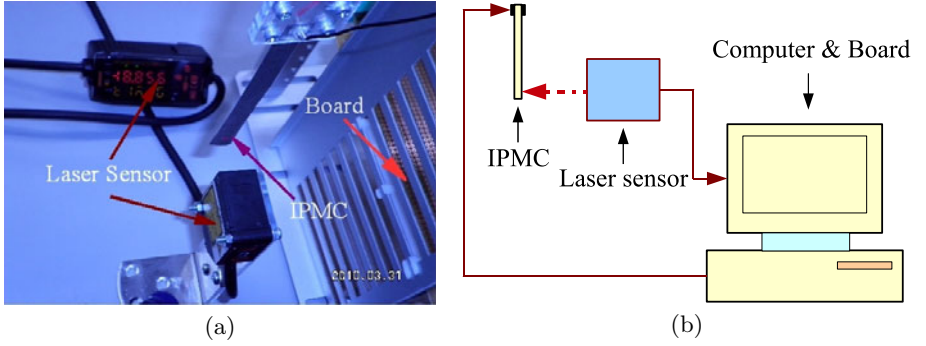
The outline of the paper is given as follows. In Section 2, experimental system is described. In Section 3, modeling and problem statement are introduced. In Section 4, robust right coprime factorization approach is presented for an IPMC set. The simulation and experimental results are shown in Section 5, and Section 6 is the conclusion.

## 2 Experimental System

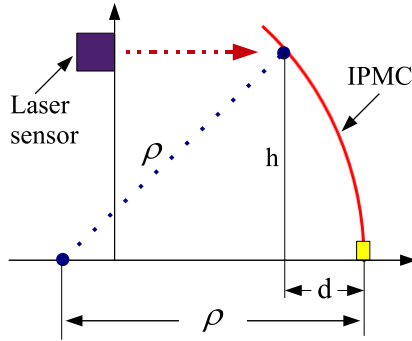
Fig. 1(a) shows picture of experimental system. The experimental system schematic illustration is shown in Fig. 1(b). In this experimental system, an IPMC sample of dimensions 50 mm 10 mm 0.2 mm is clamped at one end, and is subject to voltage excitation generated from the computer and board (PCI-3521). A laser displacement sensor (ZX-LD40: 40±10mm) is used to measure the bending displacement  $d$ . The relationship between the bending curvature  $1/\rho$  and the displacement  $d$  can be described by (see Fig. 2):

$$\frac{1}{\rho} = \frac{2d}{d^2 + h^2}. \quad (1)$$

where,  $h$  is the vertical distance.



**Fig. 1.** (a) Picture of experimental system; (b) Experimental system schematic illustration



**Fig. 2.** Calculation of the displacement  $d$  based on the curvature  $1/\rho$

### 3 Modeling and Problem Statement

#### 3.1 Nonlinear Model of IPMC

Displacement control models of IPMCs fall into two general categories: linear models, and nonlinear models. Linear models have no prior knowledge or some knowledge of the system. Nonlinear models have a comprehensive knowledge of the physics system derivation. A nonlinear dynamic model with uncertainties of IPMC can be described by [14]:

$$\begin{aligned} \dot{v} &= -\frac{v-u}{C_1(v)(R_a+R_c)}, \\ y &= \frac{3\alpha_0\kappa_e\sqrt{2\Gamma(v)}}{Y_e h^2} + \Delta P. \end{aligned} \quad (2)$$

where,  $v$  is the state variable,  $u$  is the control input voltage,  $y$  is the curvature output,  $R_a$  is the electrodes resistance,  $R_c$  is the ion diffusion resistance,  $\kappa_e$  is the effective dielectric constant of the polymer,  $\alpha_0$  is the coupling constant,  $Y_e$  is the equivalent Young's modulus,  $\Delta P$  denotes uncertainties which consist of

measurement error of parameters and model error of the IPMC.  $\Gamma(v)$  and  $C_1(v)$  are functions of the state variable and some parameters,

$$\Gamma(v) = \frac{b}{a^2} \left( \frac{ave^{-av}}{1 - e^{-av}} - \ln\left(\frac{ave^{-av}}{1 - e^{-av}}\right) - 1 \right). \quad (3)$$

here,

$$a = \frac{F}{RT}, \quad b = \frac{F^2 C^-}{RT \kappa_e}. \quad (4)$$

$F$  is Faraday's constant,  $C^-$  is the anion concentrations,  $R$  is the gas constant,  $T$  is the absolute temperature.

$$C_1(v) = S \kappa_e \frac{\dot{\Gamma}(v)}{\sqrt{2\Gamma(v)}}, \quad (5)$$

where,  $S = WL$  is the surface area of the IPMC,  $L$ ,  $W$  and  $h$  denote the length, the width and the thickness of the IPMC respectively.

According to (2), (3) and (5), the following equations are established,

$$\begin{aligned} \dot{v} &= - \frac{(v-u) \sqrt{2b \left( \frac{ave^{-av}}{1-e^{-av}} - \ln\left(\frac{ave^{-av}}{1-e^{-av}}\right) - 1 \right)}}{S \kappa_e b (R_a + R_c) \left( 1 - \frac{1-e^{-av}}{ave^{-av}} \right) \frac{e^{-av} (1-e^{-av}-av)}{(1-e^{-av})^2}} \\ y &= \frac{3\alpha_0 \kappa_e \sqrt{2b \left( \frac{ave^{-av}}{1-e^{-av}} - \ln\left(\frac{ave^{-av}}{1-e^{-av}}\right) - 1 \right)}}{a Y_e h^2} + \Delta P \end{aligned} \quad (6)$$

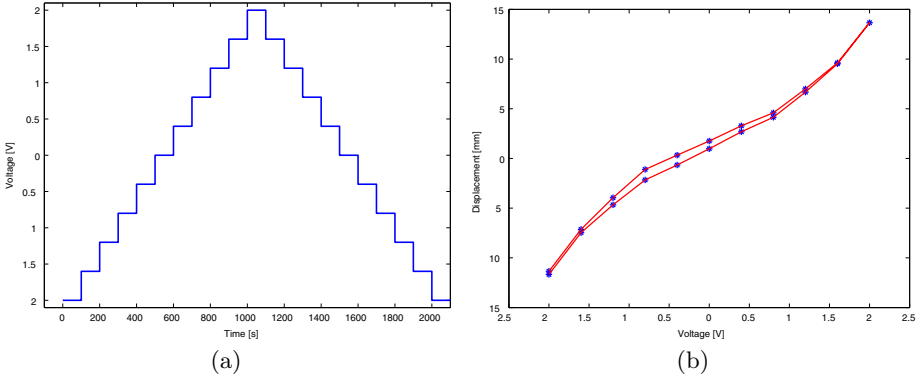
Defining a new state variable  $x = av$ , the following equations can be obtained,

$$\begin{aligned} \dot{x} &= - \frac{(x-au) \sqrt{2b \left( \frac{xe^{-x}}{1-e^{-x}} - \ln\left(\frac{xe^{-x}}{1-e^{-x}}\right) - 1 \right)}}{S \kappa_e b (R_a + R_c) \left( 1 - \frac{1-e^{-x}}{xe^{-x}} \right) \frac{e^{-x} (1-e^{-x}-x)}{(1-e^{-x})^2}} \\ y &= \frac{3\alpha_0 \kappa_e \sqrt{2b \left( \frac{xe^{-x}}{1-e^{-x}} - \ln\left(\frac{xe^{-x}}{1-e^{-x}}\right) - 1 \right)}}{a Y_e h^2} + \Delta P \end{aligned} \quad (7)$$

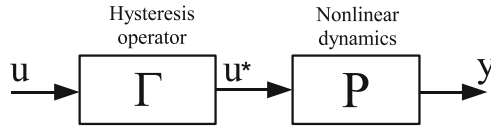
For the IPMC actuators, hysteresis is also one of the most important properties. On the existence of the hysteretic nonlinearities, the system usually exhibits undesirable oscillations and even instability. To avoid oscillations, hysteresis phenomenon must be considered.

### 3.2 Hysteresis in IPMC

The displacement response of IPMC is measured under a sequence of voltage values which are monotonically from -2.0 V to 2.0 V and then back to -2.0 V (see Fig. 3(a)). Each voltage value  $v$  is held for 100 seconds to guarantee the IPMC to reach the steady-state, which makes sure that the effects of other dynamics



**Fig. 3.** (a) The sequence of inputs applied to IPMC sample; (b) Steady-state displacement *vs* voltage input



**Fig. 4.** Model structure for an IPMC actuator

are eliminated or minimized, and any measured input-output loop would indeed come from hysteresis.

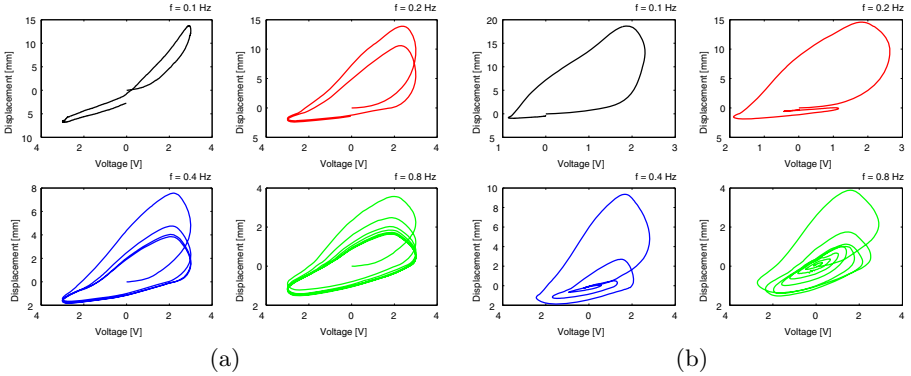
It is evident from the measured steady-state displacement *vs*. the voltage input (see Fig. 3(b)) that the input-output relationship shows hysteresis. The controlled IPMC is either nonlinear or with uncertainties and hysteresis, model structure for an IPMC actuator can be shown in Fig. 4.

### 3.3 The Prandtl-Ishlinskii Model

Prandtl-Ishlinskii hysteresis model is used to describe the hysteresis. The PI model is formulated through a weighted superposition of elementary hysteresis operators. The elementary hysteresis operators are stop hysteresis operators and play hysteresis operators [15, 16].

In this paper, PI model defined by play operators with threshold value  $h > 0$ , is used. Analytically, suppose that  $C_m[0; t_E]$ , where  $0 = t_0 < t_1 < \dots < t_N = t_E$  is a partition of  $[0; t_E]$ , is the space of piecewise monotone continuous functions. The input function  $u(t) \in C_m[0; t_E]$  is monotone on each of the sub-intervals  $[t_i; t_{i+1}]$ . The play hysteresis operator  $F_h(\cdot; u_{-1}^*) : C_m[0; t_E] \times u_{-1}^* \mapsto C_m[0; t_E]$  for an initial value  $u_{-1}^* \in R$ , is defined as

$$\begin{aligned}
 F_h[u(0); u_{-1}^*] &= f_h(u(0), u_{-1}^*) \\
 F_h[u(t); u_{-1}^*] &= f_h(u(t), F_h[u(t_i); u_{-1}^*]) \\
 &\text{for } t_i < t \leq t_{i+1} \text{ and } 0 \leq i \leq N - 1
 \end{aligned} \tag{8}$$



**Fig. 5.** The response of the displacement, (a)  $u(t) = 3\sin(2\pi ft)$ ; (b)  $u(t) = \frac{3\sin(2\pi ft)}{t+1}$

with

$$f_h(u, q) = \max(u - h, \min(u + h, q)) \tag{9}$$

The spaces of input and output can be extended to the space  $C[0; t_E]$  of continuous functions. Here the play operator is represented in another format [18].

$$F_h(u)(t) = \begin{cases} u(t) + h, & u(t) \leq F_h(u)(t_i) - h \\ F_h(u)(t_i), & -h < u(t) - F_h(u)(t_i) < h \\ u(t) - h, & u(t) \geq F_h(u)(t_i) + h \end{cases} \tag{10}$$

where  $F_h(u)$  shows  $F_h(u; u^*_{-1})$  in brief. The initial condition of (10) is given by  $F_h(u)(0) = \max(u(0) - h; \min(u(0) + h; u^*_{-1}))$ . PI model can be described by a weighted superposition of the above play operator with threshold value H.

$$u^*(t) = \Gamma(u)(t) = D_{PI}(u)(t) + \Delta(u)(t) \tag{11}$$

where

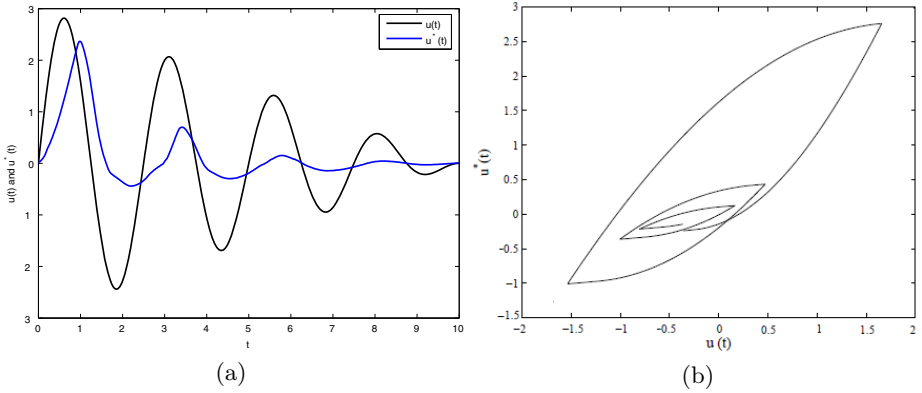
$$D_{PI}(u)(t) = Ku(t), K = \int_0^{h_x} p(h)dh$$

$$\Delta(u)(t) = - \int_0^{h_x} S_n hp(h)dh + \int_{h_x}^H p(h)F_h(u)(t_i)dh$$

$$S_n = Sgn(u(t) - F_h(u)(t_i))$$

where  $p(h)$  is a given continuous density function, satisfying  $p(h) \geq 0$  with  $\int_0^\infty hp(h)dh < \infty$ , and is expected to be identified from experimental date. To identify the density function parameters in a density function  $p(h) = a \times e^{b(h-1)^2}$ , we set the threshold value  $h$ . The weight parameters  $a$  and  $b$  are found by performing nonlinear least squares fit. To evaluate the above PI model, an experiment is conducted. In the experiment, input voltage of  $u(t) = 3\sin(2\pi ft)$  ( $f = 0.1, 0.2, 0.4, 0.8$ ) are applied to IPMC actuator. Fig. 5 shows the response of





**Fig. 6.** (a) The input and output of the identified PI hysteresis model; (b) The response of the identified PI hysteresis model

the displacement of IPMC. Where parameters  $p(h) = 0.00024 \times e^{-0.00075(h-1)^2}$  and  $h \in [0, 3]$  of PI model are found by performing a nonlinear least squares fit using Levenberg-Marquardt method. Fig. 6(a) shows the input and output of the identified PI hysteresis model. The response of the identified PI hysteresis model is shown in Fig. 6(b).

### 3.4 Problem Statement

The objective of this paper is to design a robust nonlinear control for IPMC with uncertainties and hysteresis using operator-based robust right coprime factorization so that the validity of the proposed model can be confirmed and the precision position control of IPMC can be implemented.

## 4 Design of Robust Stable Control System

Right factorization is to factorize a given plant operator  $P$  as a composition of two operators  $N$  and  $D$  in the following form

$$P = ND^{-1} \quad (12)$$

If there are two operators  $A$  and  $B$ , such that the following condition is satisfied.

$$AN + BD = I \quad (13)$$

where  $I$  is the identity operator, then the right factorization is said to be coprime [8, 9].

Consider the nonlinear system with bounded uncertainties, the robust control problem by using robust right coprime factorization approach has been researched in [10, 11, 12]. Assume that the uncertainties are included in  $\Delta P$ , where

$\Delta P$  is unknown but bounded. The right factorization of the nonlinear system is in the following form

$$\tilde{P} = P + \Delta P = (N + \Delta N)D^{-1} \tag{14}$$

From [10], we can see that if the following conditions are satisfied,

$$AN + BD = L \tag{15}$$

$$A(N + \Delta N) + BD = \tilde{L} \tag{16}$$

$$\|(A(N + \Delta N) - AN)L^{-1}\| < 1 \tag{17}$$

then the stability of the uncertain system  $\tilde{P}$  is guaranteed, where  $L$  and  $\tilde{L}$  are unimodular operators and  $\|\cdot\|$  is Lipschitz operator norm. Then, we consider the mentioned nonlinear IPMC by using robust right coprime factorization. Consider hysteresis in model structure, the nonlinear dynamics can be described,

$$\begin{aligned} \dot{x} &= -\frac{(x - au^*)\sqrt{2b\left(\frac{xe^{-x}}{1-e^{-x}} - \ln\left(\frac{xe^{-x}}{1-e^{-x}}\right) - 1\right)}}{S\kappa_e b(R_a + R_c)\left(1 - \frac{1-e^{-x}}{xe^{-x}}\right)\frac{e^{-x}(1-e^{-x}-x)}{(1-e^{-x})^2}} \\ y &= \frac{3\alpha_0\kappa_e\sqrt{2b\left(\frac{xe^{-x}}{1-e^{-x}} - \ln\left(\frac{xe^{-x}}{1-e^{-x}}\right) - 1\right)}}{aY_e h^2} + \Delta P \end{aligned} \tag{18}$$

$D$ ,  $N$  and  $\Delta N$  are denoted as the following forms

$$D(\omega)(t) = \frac{S\kappa_e b(R_a + R_c)\dot{\omega}(t)\left(1 - \frac{1-e^{-\omega(t)}}{\omega(t)e^{-\omega(t)}}\right)\frac{e^{-\omega(t)}(1-e^{-\omega(t)}-\omega(t))}{(1-e^{-\omega(t)})^2}}{a\sqrt{2b\left(\frac{\omega(t)e^{-\omega(t)}}{1-e^{-\omega(t)}} - \ln\left(\frac{\omega(t)e^{-\omega(t)}}{1-e^{-\omega(t)}}\right) - 1\right)}} + \frac{\omega(t)}{a} \tag{19}$$

$$N(\omega)(t) = \frac{3\alpha_0\kappa_e\sqrt{2b\left(\frac{\omega(t)e^{-\omega(t)}}{1-e^{-\omega(t)}} - \ln\left(\frac{\omega(t)e^{-\omega(t)}}{1-e^{-\omega(t)}}\right) - 1\right)}}{aY_e h^2} \tag{20}$$

$$\Delta N(\omega)(t) = \Delta\sqrt{2b\left(\frac{\omega(t)e^{-\omega(t)}}{1-e^{-\omega(t)}} - \ln\left(\frac{\omega(t)e^{-\omega(t)}}{1-e^{-\omega(t)}}\right) - 1\right)} \tag{21}$$

where  $\Delta N$  is the numerator of the coprime factorization presentation of  $\Delta P$ ,  $N$ ,  $\Delta N$ , and  $D$  are stable operators respectively, and  $D$  is invertible.

For the IPMC, based on robust right coprime factorization method, a robust stable control for a nonlinear model is designed. The framework of control system is shown in Fig. 7.  $U$  is the input space of PI model, and  $U^*$  is the output spaces of PI model. Let the output space of the original nonlinear plant and quasi-state space be  $Y$  and  $W$ .  $N$ ,  $\Delta N$ ,  $D$  are  $N : W \rightarrow Y$ ,  $\Delta N : W \rightarrow Y$ ,  $D : W \rightarrow U^*$ , respectively.  $A$ ,  $B$  are the stable operator controllers and  $B$  is invertible. We can choose  $W = U$ . In Fig. 7, the output  $u^*$  of PI model is the input of the original nonlinear plant. In practice, the signal of  $u^*$  between the hysteresis

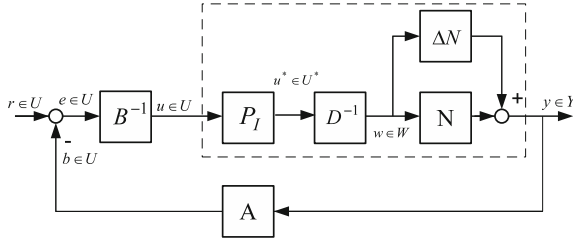


Fig. 7. A robust control system with uncertainties and hysteresis

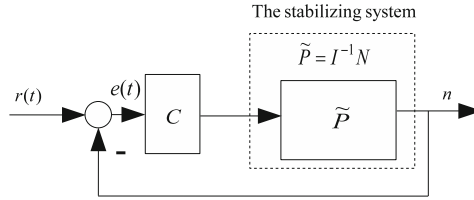


Fig. 8. The tracking control system

and the nonlinear plant is unavailable. So the hysteresis of actuator should be considered as a part of the nonlinear plant during designing controllers. We consider  $D_{PI}(u)(t)$  in the hysteresis model given in (11) and  $D$  of the above plant as a new invertible factor of right coprime factorization  $\tilde{D}(\omega)(t) = D_{PI}^{-1}D(\omega)(t)$ . The corresponding controllers  $A$  and  $B$  which satisfy Bezout identity  $AN + B\tilde{D} = M \in S(W, U)$  are also designed to make the nonlinear system be BIBO stable, where  $S(W, U)$  is the set of unimodular operator [15, 16].

Based on the proposed design scheme, we design controllers as follows.

$$A(y(t)) = -\frac{aSY_e h^2(R_a + R_c)}{3\alpha_0} \dot{y}(t) \tag{22}$$

$$B(u(t)) = aKu(t) \tag{23}$$

where  $K = \int_0^{h_x} p(h)dh$ .

Besides the robust stability of the IPMC position control system is guaranteed, the tracking performance of the system is also considered. Here, the tracking condition is difficult to obtain because the operator  $N$  is a complex nonlinear function, therefore we design a tracking system shown in Fig. 8, where the stabilizing system regarded as the plant is equal to the system in Fig. 7.

The tracking controller  $C$  is designed satisfying the conditions in [11]. Here, the controller  $C$  is designed as

$$u(t) = k_p e(t) + k_i \int e(\tau) d\tau \tag{24}$$

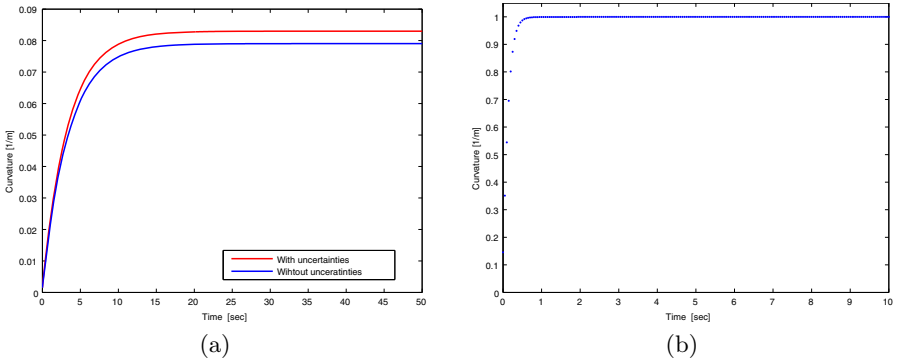
## 5 Simulation and Experimental Results

In this section, some simulation and experimental results are given to illustrate the effectiveness of the proposed method. Some physical parameters have been identified through experiments, which are shown in Table 1.

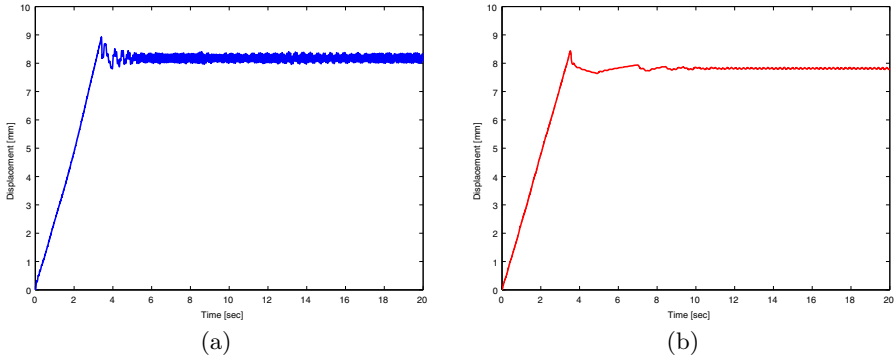
**Table 1.** Parameters in the model

T	F	$\kappa_e$
290 K	96487 C mol <sup>-1</sup>	$1.12 \times 10^{-6} Fm^{-1}$
$R_a$	R	$Y_e$
18 $\Omega$	8.3143 J mol <sup>-1</sup> K <sup>-1</sup>	0.56 GPa
$R_c$	$C^-$	$\alpha_0$
60 $\Omega$	980 mol	0.12 J C <sup>-1</sup>
L	W	h
50 mm	10 mm	200 $\mu$ m

During the simulations and experiment, the uncertain factor in (21) is modeled as  $\Delta = \frac{3\alpha_0\kappa_e\sqrt{2b}}{aY_e h^2} \times 5\%$ , the uncertainties of model are smaller than  $\Delta N$  in practice. The simulation result of the IPMC with uncertainties and without uncertainties of control system based on right coprime factorization is shown in Fig. 9(a). Fig. 9(b) shows the simulation result of system with tracking controller where the reference input of the curvature is  $r_f = 1[1/m]$ . From Fig. 9, we can find that the nonlinear IPMC with uncertainties control system using right coprime factorization control is robust stable. By the tracking controller, the IPMC output can track the reference input. Fig. 10 shows the displacement response of the IPMC using robust right coprime factorization, the desired outputs of displacement  $d$  is 8 [mm]. These results show that the robust stability of the IPMC displacement control system is guaranteed and tracking performance is



**Fig. 9.** (a) The simulation result based on robust coprime factorization; (b) The simulation result with tracking controller



**Fig. 10.** Experimental results on output response, (a) Without PI model; (b) With PI model

satisfied by using the proposed method. The oscillations in the system responses can be avoided by considering hysteresis of IPMC actuator.

## 6 Conclusions

In this paper, we investigate robust nonlinear control for the ionic polymer metal composite with hysteresis and uncertainties. The nonlinear model of the IPMC with uncertainties and hysteresis is obtained. The Prandtl-Ishlinskii model is used to model the hysteresis. To avoid the influence of the uncertainties and hysteresis, robust right coprime factorization approach for the IPMC displacement control system is proposed. Finally, simulation and experimental results are presented to show the effectiveness of the proposed method.

## References

1. Shahinpoor, M., Kim, K.: Ionic polymer-metal composites: I. Fundamentals. *Smart Materials and Structures* 10, 819–833 (2001)
2. Shahinpoor, M.: Ionic polymer-conductor composites as biomimetic sensors, robotic actuators and artificial muscles—a review. *Electrochimica Acta* 48, 2343–2353 (2003)
3. Chen, Z., Tan, X.: A control-oriented and physics-based model for ionic polymer-metal composite actuators. *IEEE/ASME Transactions on Mechatronics* 13, 519–529 (2008)
4. Bhat, N., Kim, W.: Precision position control of ionic polymer metal composite. In: *Proc. of the 2004 American Control Conference*, Boston, MA, pp. 740–745 (2004)
5. Oh, S., Kim, H.: A study on the control of an IPMC actuator using an adaptive fuzzy algorithm. *Journal of Mechanical Science and Technology* 18, 1–11 (2004)
6. Bhat, N., Kim, W.: Precision force and position control of ionic polymer metal composite. *Journal of Systems and Control Engineering* 218, 421–432 (2004)
7. Hao, L., Li, Z.: Modeling and adaptive inverse control of hysteresis and creep in ionic polymer-metal composite actuators. *Smart Materials and Structures* 19, 14–25 (2010)

8. Deng, M., Inoue, A., Ishikawa, K.: Tracking of perturbed nonlinear plants using robust right coprime factorization approach. In: Proc. of the 2004 American Control Conference, Boston, MA, pp. 3666–3670 (2004)
9. Figueiredo, R., Chen, G.: Nonlinear Feedback Control System: an Operator Theory Approach. Academic Press Inc., New York (1993)
10. Deng, M., Inoue, A., Ishikawa, K.: Operator-based nonlinear feedback control design using robust right coprime factorization. *IEEE Transactions on Automatic Control* 51, 645–648 (2006)
11. Chen, G., Han, Z.: Robust right coprime factorization and robust stabilization of nonlinear feedback control systems. *IEEE Transactions on Automatic Control* 43, 1505–1510 (1998)
12. Deng, M., Bi, S., Inoue, A.: Robust nonlinear control and tracking design for multi-input multi-output nonlinear perturbed plants. *IET Control Theory & Applications* 3, 1237–1248 (2009)
13. Wen, S., Deng, M.: Experimental study of robust control to a nonlinear MIMO process using robust right coprime factorization. *ICIC Express Letters* 3, 1061–1066 (2009)
14. Deng, M., Wang, A., Minami, M., Yanou, A.: Operator-based Modeling for Nonlinear Ionic Polymer Metal Composite with Uncertainties. In: Proc. of 5th International Conference on Soft Computing and Intelligent Systems and 11th International Symposium on Advanced Intelligent Systems, Okayama, Japan (2010) (submitted)
15. Saito, S., Deng, M., Inoue, A., Jiang, C.: Vibration control of a flexible arm experimental system with hysteresis of piezoelectric actuator. *International Journal of Innovative Computing, Information and Control* 6, 2965–2975 (2010)
16. Jiang, C., Deng, M., Inoue, A.: Robust stability of nonlinear plants with a non-symmetric Prandtl-Ishlinskii hysteresis model. *International Journal of Automation and Computing* 7, 213–218 (2010)

# Modeling of Hysteresis Nonlinearity Based on Generalized Bouc-Wen Model for GMA

Yongxin Guo and Jianqin Mao

School of Automation Science and Electrical Engineering, Beihang University,  
Beijing 100191, China  
gyx1979\_1999@yahoo.com.cn

**Abstract.** In this paper, the classical and generalized Bouc-Wen models are used to describe the hysteresis loops of Giant Magnetostrictive Actuator (GMA) for the first time. The parameters of the models are optimized with Particle Swarm optimization (PSO) algorithm. The hysteresis loops of GMA are asymmetric, and the classical Bouc-Wen model can not describe such asymmetric hysteresis loops satisfactorily, but the generalized BoucWen model can describe asymmetric hysteresis loops better. Simulation results prove the correctness of the conclusion.

**Keywords:** GMA; hysteresis; asymmetric; generalized Bouc-Wen model; Particle Swarm Optimization.

## 1 Introduction

Giant Magnetostrictive Material (GMM) is a type of functional material. It can produce stress and strain when a magnetic field is applied to it. Compared to conventional magnetostrictive and piezoelectric materials, it has higher strain value, more rapid response speed, better frequency characteristics and higher reliability. Furthermore, it is not subject to fatigue, overheating and failure. GMM is widely used in the areas of aviation, aerospace, and precise manufacturing, etc. One of the most important applications is to construct actuators which are often called Giant Magnetostrictive Actuators (GMA). A type of giant magnetostrictive actuator produced by Department of Materials Science and Engineering of Beihang University is depicted in Fig. 1.

However, hysteresis nonlinearity is the inherent property of GMA, it can reduce the control accuracy and even lead to oscillation. There are mainly three classes of approaches for the modeling of hysteresis. One is the physical modeling theory, which is represented by Jiles-Atherton model [1] and Duhem model [2]. The other is the operator modeling theory, which is represented by Preisach model [3] Krasnosel'skii-Pokrovskii (KP) model[4], Prandtl-Ishlinskii (PI) model [5], etc. Another is the intelligent modeling theory based on computational intelligence [6]. All the three classes are widely used.

The Bouc-Wen model is a semi-physical model which was initially proposed by R.Bouc in 1971 and generalized by Y.K.Wen in 1976. The model has the advantage of computational simplicity, because it needs only one auxiliary nonlinear differential equation to describe the hysteretic behaviour. It is widely used in the current

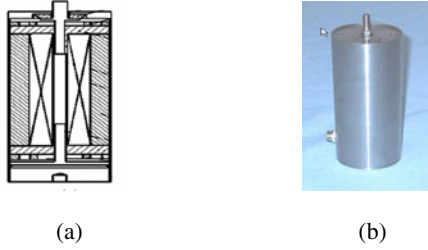


Fig. 1. Giant Magnetostrictive Actuator. (a) Section. (b) Prototype.

literature, particularly within areas of civil and mechanical engineering. However, according to the published papers, few people have used the Bouc-Wen model to describe the hysteresis loops of GMA.

In this paper, the hysteresis loops of GMA is modeled with the classical and the generalized Bouc-Wen model *for the first time* and the parameters of the models are optimized with the Particle Swarm Optimization (PSO) algorithm.

This paper is organized as follows: In section 2, the Bouc-Wen model is briefly introduced. In section 3, approaches of modeling the hysteresis of GMA are presented and parameters of the models are optimized with PSO algorithm. In section 4, simulation is carried out and the results are discussed. Finally, some concluding remarks are stated in section 5.

## 2 Bouc-Wen Model

In Ref. [7], the Bouc-Wen model takes the form:

$$\begin{aligned} \Phi_{BW}(x)(t) &= \alpha kx(t) + (1 - \alpha)Dkz(t), \\ \dot{z} &= D^{-1}(A\dot{x} - \beta|\dot{x}|z|z|^{n-1} - \gamma\dot{x}|x|^n) . \end{aligned} \tag{1}$$

Where the over-dot denotes the derivative with respect to time and  $0 < \alpha < 1$ ,  $k > 0$ ,  $D > 0$ ,  $n > 1$ ,  $\beta + \gamma \neq 0$ . Equation (1) is called the standard Bouc-Wen model. The normalized version of the Bouc-Wen model, takes the form:

$$\begin{aligned} \Phi_{BW}(x)(t) &= \kappa_x x(t) + \kappa_w w(t), \\ \dot{w}(t) &= \rho(\dot{x} - \sigma|\dot{x}(t)||w(t)|^{n-1} w(t) + (\sigma - 1)\dot{x}(t)|w(t)|^n) . \end{aligned} \tag{2}$$

Where

$$\begin{aligned} w(t) &= \frac{z(t)}{z_0}, z_0 = \sqrt[n]{\frac{A}{\beta + \gamma}}, \rho = \frac{A}{Dz_0}, \sigma = -\frac{\beta}{\beta + \gamma}, \\ \kappa_x &= \alpha k > 0, \kappa_w = (1 - \alpha)Dkz_0 > 0 . \end{aligned} \tag{3}$$

The standard and normalized form are called classical Bouc-Wen model here. The output of the classical Bouc-Wen is a *point symmetric* loop, as depicted in Fig. 2. However, in practice, the hysteresis loops may exhibit asymmetric shape due to asymmetry in geometric bound conditions or material properties, such as the



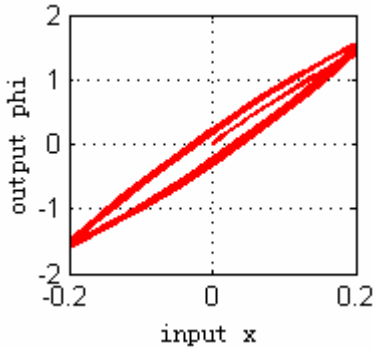


Fig. 2. Loops of classical Bouc-Wen model

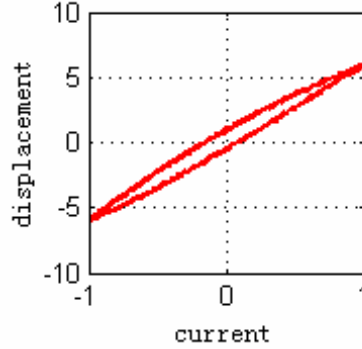


Fig. 3. Asymmetric loops of GMA

hysteresis loops of GMA. Fig. 3 depicts the hysteresis loop of a GMA which bulges upward (the input electric current is 1A, 0.5Hz). So, modeling asymmetric hysteresis loops of GMA with classical Bouc-Wen model may lead to large error.

To account for the asymmetry, Junho Song and Armen Der Kiureghian introduced a generalized Bouc-Wen model [8] which has enhanced flexibility in shape control. It can describe highly asymmetric hysteresis with fixed parameters. The generalized Bouc-Wen model takes the form:

$$\begin{aligned}
 f_s(x, \dot{x}, z) &= \alpha k_0 x + (1 - \alpha) k_0 z, \\
 \dot{z} &= \dot{x} [A - |z|^n \psi(x, \dot{x}, z)] .
 \end{aligned}
 \tag{4}$$

Where  $\dot{z}$  is the derivative of  $z$  with respect to time,  $A$  and  $n$  are parameters that control the scale and sharpness of the hysteresis loop respectively,  $\psi(x, \dot{x}, z)$  is the nonlinear function of  $x, \dot{x}, z$  that controls other shape features of the hysteresis loop. The function  $\psi(x, \dot{x}, z)$  is called shape control function and it could take different forms. The form used in Ref. [8] is:

$$\psi = \beta_1 \operatorname{sgn}(\dot{x}z) + \beta_2 \operatorname{sgn}(x\dot{x}) + \beta_3 \operatorname{sgn}(xz) + \beta_4 \operatorname{sgn}(\dot{x}) + \beta_5 \operatorname{sgn}(z) + \beta_6 \operatorname{sgn}(x) .
 \tag{5}$$

Where  $\beta_1, \dots, \beta_6$  are fixed parameters and  $\operatorname{sgn}()$  is the sign function. This model has six degrees of freedom as it can control the values of  $\psi(x, \dot{x}, z)$  at six phases. The six phases are inferred as follows:

phase1:  $x > 0, \dot{x} > 0, z > 0$ , where  $\psi(x, \dot{x}, z) = \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6$ ,

phase2:  $x > 0, \dot{x} < 0, z > 0$ , where  $\psi(x, \dot{x}, z) = -\beta_1 - \beta_2 + \beta_3 - \beta_4 + \beta_5 + \beta_6$ ,

phase3:  $x > 0, \dot{x} < 0, z < 0$ , where  $\psi(x, \dot{x}, z) = \beta_1 - \beta_2 - \beta_3 - \beta_4 - \beta_5 + \beta_6$ ,

phase4:  $x < 0, \dot{x} < 0, z < 0$ , where  $\psi(x, \dot{x}, z) = \beta_1 + \beta_2 + \beta_3 - \beta_4 - \beta_5 - \beta_6$ ,

phase5:  $x < 0, \dot{x} > 0, z < 0$ , where  $\psi(x, \dot{x}, z) = -\beta_1 - \beta_2 + \beta_3 + \beta_4 - \beta_5 - \beta_6$ ,

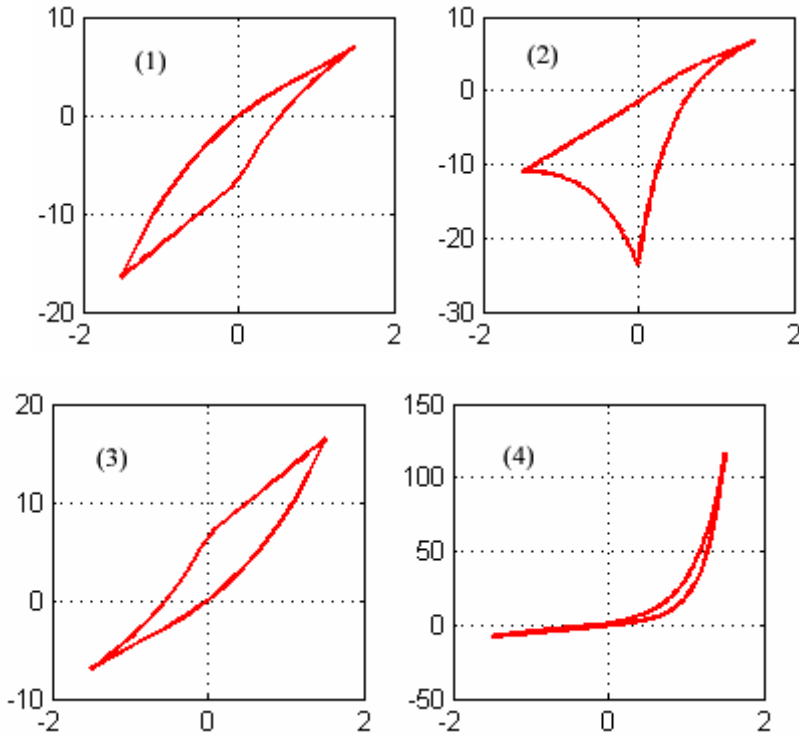
phase6:  $x < 0, \dot{x} > 0, z > 0$ , where  $\psi(x, \dot{x}, z) = \beta_1 - \beta_2 - \beta_3 + \beta_4 + \beta_5 - \beta_6$ .

With different combinations of the parameters of  $\beta_1, \dots, \beta_6$  as listed in table 1 ( $k_0=8, \alpha = 0.8, A=2, n=1$ ), the generalized Bouc-Wen model can describe loops of asymmetric shapes, as depicted in Fig.4.

**Table 1.** Parameters of the Generalized Bouc-Wen Model

	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$
(1)	1	1	1	1	1	1
(2)	1	1	3	1	1	1
(3)	1	1	-1	-1	-1	1
(4)	1	1	-3	-1	-1	-1

With equation (4) and (5), the parameters of the generalized Bouc-Wen model can be adjusted to minimize the difference between the output of the model and the experimental data over the entire hysteresis loop of GMA.



**Fig. 4.** Loops of the generalized Bouc-Wen model with different parameters

### 3 Modeling and Parameter Optimization

#### 3.1 Discrete Generalized Bouc-Wen Model

From equation (4) and (5), the discrete form of the generalized Bouc-Wen model is expressed as (first order forward difference):

$$\begin{aligned} \hat{y}(k+1) &= \alpha K_0 x(k) + (1-\alpha)K_0 z(k), \\ z(k+1) &= (x(k+1) - x(k))(A - |z(k)|^n) \psi + z(k) \end{aligned} \tag{6}$$

$$\begin{aligned} \psi &= \beta_1 \operatorname{sgn}((x(k+1) - x(k))z(k)) + \beta_2 \operatorname{sgn}(x(k)(x(k+1) - x(k))) + \\ &\beta_3 \operatorname{sgn}(x(k)z(k)) + \beta_4 \operatorname{sgn}(x(k+1) - x(k)) + \beta_5 \operatorname{sgn}(z(k)) + \beta_6 \operatorname{sgn}(x(k)) \end{aligned} \tag{7}$$

where  $x(k)$  and  $y(k)$  are the input and output of the actuator at time  $k$  and  $\hat{y}(k+1)$  is the output of the model at time  $k+1$ . The model essentially predicts the output of the actuator at time  $k+1$  using the current and historic inputs and outputs of the actuator. (Discretization of the normalized Bouc-Wen model is similar to that of the generalized one).

#### 3.2 Particle Swarm Optimization (PSO) Algorithm [9]

Due to its highly nonlinear nature, parameter identification of Bouc-Wen systems constitutes a challenging problem. PSO algorithm is a kind of evolutionary algorithm developed in recent years, which has been shown to be simple and effective in parameter optimization. It is deployed in this paper to identify the parameters of the Bouc-Wen model.

PSO algorithm treats each set of parameters as one particle. The number of the parameters is the particle’s dimension. Each particle knows its current position and the past best position. Moreover, each particle knows the past best position of the group. Each particle moves to a new position according to the previous positions and the new velocity of the particle. The new position is compared with the best position generated by previous particles and the best one is kept. If a particle discovers a new probable solution, other particles will move closer to it to explore the region more completely.

There are three attributes for each particle: current position (current value of the particle), current velocity (the rate of the position change) and past best position. The new velocity and position of each particle can be calculated by the formulas (8), (9) and (10). Each particle  $i$  in the  $d$ th sub-dimensional space is iteratively updated until the best solution is gained or the iteration reaches the allowable maximum number.

$$v_{id}^{t+1} = wv_{id}^t + c_1 \operatorname{rand}_1() (p_{id}^t - x_{id}^t) + c_2 \operatorname{rand}_2() (p_{gd}^t - x_{id}^t), \tag{8}$$

$$\begin{cases} v_{id}^t = v_d^{\max}, & \text{if } v_{id}^t > v_d^{\max}, \\ v_{id}^t = -v_d^{\max}, & \text{if } v_{id}^t < -v_d^{\max}. \end{cases} \tag{9}$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1}, \tag{10}$$

$$i = 1, 2, \dots, N; d = 1, 2, \dots, D.$$

where  $v_{id}^t$  is the velocity of particle  $i$  in the  $d$ th sub-dimensional space at iteration  $t$ ;  $w$  is the inertia weighting factor;  $c_1$  and  $c_2$  are the acceleration constants;  $rand_1()$  and  $rand_2()$  are random numbers between 0 and 1;  $p_{id}^t$  is the past best position of particle  $i$  in the  $d$ th sub-dimensional space at iteration  $t$ ;  $p_{gd}^t$  is the past best position of the group in the  $d$ th sub-dimensional space at iteration  $t$ ;  $v_d^{\max}$  is the maximum velocity of the particle in the  $d$ th sub-dimensional space;  $x_{id}^t$  is the position of particle  $i$  in the  $d$ th sub-dimensional space at iteration  $t$ ;  $N$  is the number of particles;  $D$  is the dimension of each particle.

There are two key points for the PSO algorithm: the coding of the solution and the fitness function. The PSO algorithm uses real-value encoding, the initialization of the algorithm can be of arbitrary values. Root-Mean-Square-Error (RMSE, in formula (11)) of Bouc-Wen model’s outputs and experiment outputs is adopted as the fitness function. The best solution is gained when the expected RMSE (ERMSE) reaches the preset value.

$$RMSE = \sqrt{\sum_{i=1}^n (f(x_i) - y_i)^2 / n}. \tag{11}$$

$$RE = \sqrt{\sum_{i=1}^n |f(x_i) - y_i|^2 / \sum_{i=1}^n |y_i|^2}. \tag{12}$$

In formula (11) and (12),  $f(x_i)$  is the output of the model.

### 3.3 Process of Modeling

Step 1: Data collection. Choose two sets of experimental data pairs  $(x(k), y(k))$  of GMA for model training and testing.

Step 2: Initialization. All the parameters  $(k_0, \alpha, A, n, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)$  of the generalized Bouc-Wen model are combined as one particle (for the normalized Bouc-Wen model, the combination is  $(\kappa_x, \kappa_w, \rho, \sigma, n)$ ). Determine the feasible ranges of the parameters and the expected RMSE (ERMSE). The feasible ranges can be big enough at first to ensure that the optimal values be in the range. Determine the number of the particles  $N$ , the range of inertial weighting factor  $[w_{\min}, w_{\max}]$ , the range of the velocity  $[v_d^{\min}, v_d^{\max}]$ , the acceleration constants  $c_1$  and  $c_2$ , and the number of iterations  $M$ . Generate the initial positions of the particles in the feasible ranges randomly.

Step 3: Evaluation. The outputs of the model are calculated by equations (6) and (7). Take formula (11) as the fitness function. Evaluate the fitness value of each particle and find the best position of each article and the group. If the value of the fitness function  $RMSE > ERMSE$  and the iteration number  $< M$ , go to step 4; otherwise go to step 5. Step 4: Position Change. Calculate the velocity and position of each particle by the formulas (8), (9) and (10). The inertial weight factor is adjusted by formula (13). Then go to step 3.

$$w = w_{\max} - (k-1)(w_{\max} - w_{\min}) / M, k = 2, \dots, M. \quad (13)$$

Step 5: Termination. The algorithm ends and the present best position of the group is the final solution.

Step 6: Test the model with the training and testing data pairs to obtain the RMSE and the RE by formula (11) and (12).

## 4 Simulation

All the data are from the experiment done by Kang Xu on April 14th, 2008. The GMA used in experiment is manufactured by Department of Materials Science and Engineering of BeiHang University. The GMA is driven by a power amplifier (GF-20). The dSPACE system is DS1103. The output displacement is measured by the eddy current sensor (the resolution is  $8mV / \mu m$ ). The sample frequency is 10 kHz. The experiment equipments are described in Fig.5.



Fig. 5. The experiment equipments

The input is quasi-static sine signal of 0.5Hz, the driving currents are 0.5A, 1A and 1.5A respectively. 5000 input-output data pairs of  $(x^i, y^i)$  are chosen as training data, and another set of 5000 pairs as testing data. The hysteresis modeling and parameters optimization is carried out according to section 3 with the training data. The number of the particles  $N$  is set as 30. The range of inertial weighting factor  $[w_{\min}, w_{\max}]$  is set as  $[0.3, 1.2]$ . Iteration number  $M$  is set as 100. Acceleration constant  $c_1$  is set as 2,  $c_2$  as 2. Then the model with the optimized parameters is tested with the

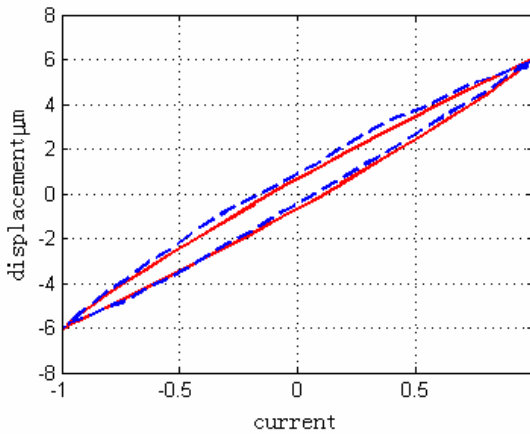
**Table 2.** Parameters of the Generalized Bouc-Wen Model

current	$k_0$	$\alpha_0$	A	n	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$
0.5A	2.70	0.92	3	3	1.2564	0.5225	-0.5403	-0.6818	0.4652	0.0630
1A	6.0	0.84	2	2	1.6350	0.6350	-0.5425	-1.4387	0.8001	0.5804
1.5A	8.99	0.77	2	1.45	1.2850	-0.0838	-0.1320	-0.6642	-0.2867	1.0305

testing data pairs. The optimized parameters of the normalized Bouc-Wen model are as follows:

$\kappa_x = 5.095, \kappa_w = 1, \rho = 1.4686, n = 1.9132, \sigma = 1.9$  (The input current is 1 A). The optimized parameters of the generalized Bouc-Wen model are listed in table 2.

The simulation result with the normalized Bouc-Wen model is depicted in Fig.6. It can be seen that the output of the normalized Bouc-Wen model doesn't coincide with the loop of GMA properly. The RMSE and RE of the normalized and generalized Bouc-Wen model are listed in table 3. It can be seen that the generalized Bouc-Wen model is more suitable to describe the hysteresis loops of GMA.



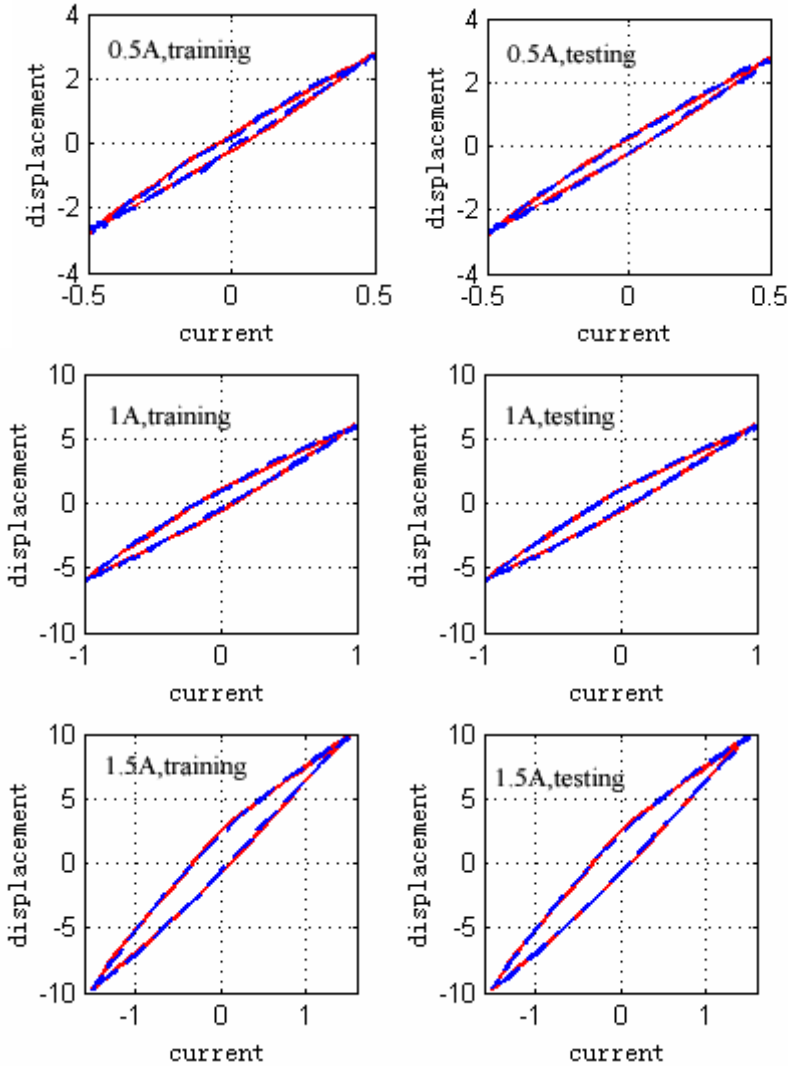
**Fig. 6.** Modeling with normalized Bouc-Wen model (1 A). Solid line: simulation ; Dashed line: experimental measurement

**Table 3.** RMSE and RE of normalized and generalized Bouc-Wen model (1 A)

Bouc-Wen	Training RMSE	Training RE	Testing RMSE	Testing RE
normalized	0.1893	0.0444	0.1851	0.0434
generalized	0.1102	0.0259	0.1071	0.0251

**Table 4.** RMSE and RE of training and testing with different currents

current	Training RMSE	Training RE	Testing RMSE	Testing RE
0.5A	0.0709	0.0365	0.0710	0.0370
1A	0.1102	0.0259	0.1071	0.0251
1.5A	0.1384	0.0199	0.1377	0.0198



**Fig. 7.** Modeling with generalized Bouc-Wen model at different currents. Solid: simulation ; Dashed: experimental measurement.

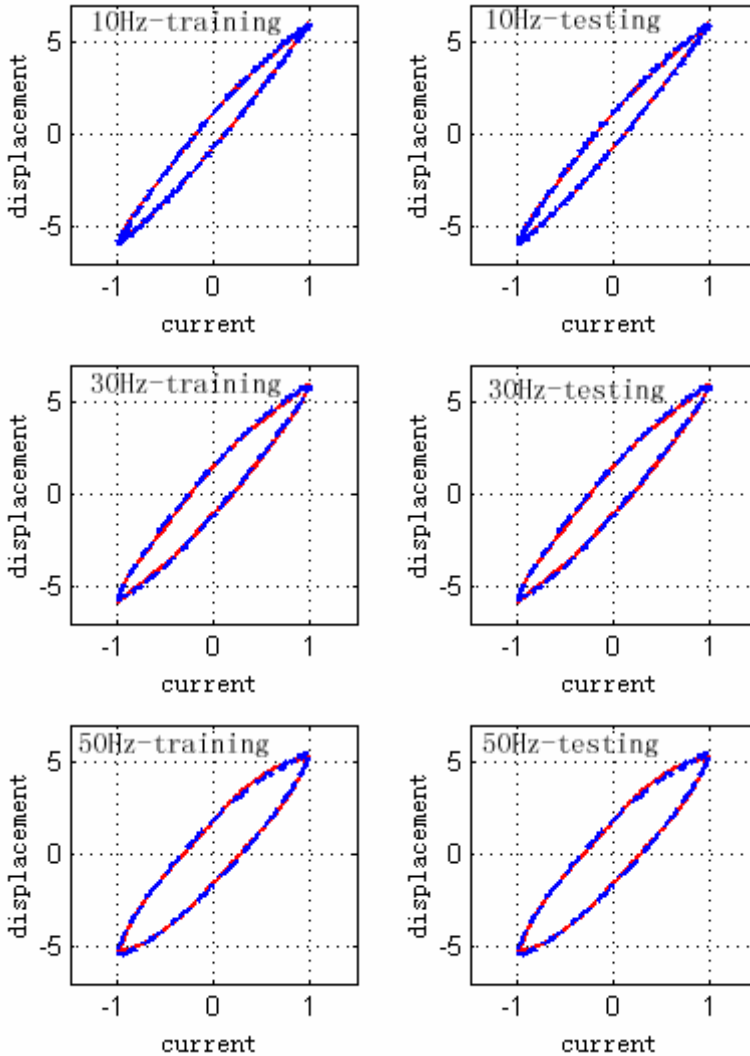
The RMSE and RE of the generalized Bouc-Wen model with different currents are listed in Table 4. The simulation results are depicted in Fig. 7.

The RMSE and RE show that the generalized Bouc-Wen model can be used to model the hysteresis loops of GMA at low frequencies.

The hysteresis nonlinearity of GMA is rate-dependent, that is to say, for inputs with different frequency, shapes of hysteresis loops are different. The classical Bouc-Wen model is rate-independent [7], so is the generalized Bouc-Wen model.

Therefore, parameters of the generalized Bouc-Wen model need to be re-identified when input frequency changes.

Fig.8 shows the modeling of hysteresis loops of GMA at different frequencies, the RMSE and RE are listed in table 5. It can be seen that the simulation result is satisfactory below 50 Hz. The RMSE and RE get larger when frequency gets higher (above 50Hz). This is because the loops of GMA show ellipse-like shapes and become more asymmetric when the frequency gets higher. The Bouc-Wen model needs to be further modified to describe such hysteresis loops of GMA.



**Fig. 8.** Modeling with generalized Bouc-Wen model at different frequencies. Solid line: simulation; Dashed line: experimental measurement.



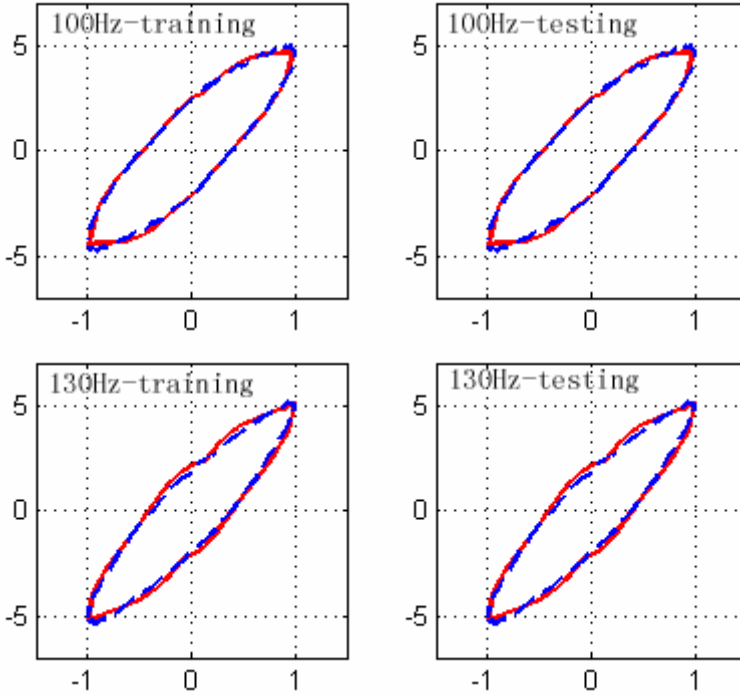


Fig. 8. (continued)

Table 5. RMSE and RE of training and testing at different frequencies

Frequency	Training RMSE	Training RE	Testing RMSE	Testing RE
10Hz	0.0991	0.0235	0.0930	0.0220
30Hz	0.1340	0.0324	0.1371	0.0332
50Hz	0.1316	0.0338	0.1276	0.0325
100Hz	0.1723	0.0494	0.1862	0.0535
130Hz	0.2213	0.0585	0.2327	0.0710

## 5 Conclusion

In this paper, a method of modeling the hysteresis nonlinearity of GMA by using the generalized Bouc-Wen model is proposed, and the parameters of the model is optimized with the Particle Swarm Optimization (PSO) algorithm. Simulation results show that this model is more suitable to describe the hysteresis loops of GMA than the normalized Bouc-Wen model at low frequencies. The range of the parameters of the generalized Bouc-Wen model is determined based on experiential judgment about the relationship between the shape of the loops and the values of the parameters. As the frequency gets higher, the Bouc-Wen model needs to be further modified to describe hysteresis loops of GMA. *The result of this paper can be considered as a stepping stone to be used towards the development of modeling the hysteresis nonlinearity of GMA with the Bouc-Wen model.*

## Acknowledgement

The authors would like to thank the anonymous reviewers for their useful comments. And this work is supported by Natural Science Foundation of China under Grant 60534020, Research Foundation for Doctoral program of China under Grant 20070006060, the Key Subject Foundation of Beijing under Grant XK100060526 and Research on Control Principles of Operation Stabilization and Maneuvering Ability for Hypersonic Vehicles Based on a Class of Smart Structure supported by the Natural Science Foundation of China.

## References

1. Jiles, D.C., Atherton, D.L.: Theory of ferromagnetic hysteresis. *Journal of Magnetism and Materials* 61, 48–60 (1986)
2. Brokate, M., Sprekels, J.: *Hysteresis and Phase Transitions*. Springer, New York (1996)
3. Mayergoyz, I.D.: Dynamic preisach model for hysteresis. *IEEE Trans. Magn.* 24(6) (1986)
4. Webb, G.V., Lagoudas, D.C., Kurdila, A.J.: Hysteresis Modeling of SMA Actuators for Control Application. *Journal of Intelligent Material Systems and Structures* 9, 432–448 (1998)
5. Kuhnen, K.: Modeling, Identification and Compensation of Complex Hysteresis Nonlinearities, A Modified Prandtl-Ishlinskii Approach. *European Journal of Control* 9(4), 407–418 (2003)
6. Mao, J., Ding, H.: Intelligent Modeling and Control for Nonlinear Systems with Rate-Dependent Hysteresis. *Science in China Press* 52(4), 547–722 (2009)
7. Ikhouan, F., Rodellar, J.: *Systems with Hysteresis- Analysis, Identification and Control using the Bouc–Wen Model*. John Wiley & Sons Ltd., England (2007)
8. Song, J., Der Kiureghian, A.: Generalized Bouc–Wen Model for Highly Asymmetric Hysteresis. *Journal of Engineering Mechanics-ASCE* 132(6), 610–618 (2006)
9. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948. IEEE, Piscataway (1995)

# Motion Dynamics Modelling of an Electric Wheel Robot<sup>\*</sup>

Xuezhu Wang, Xiangtao Zhuan, Guilin Zheng, and Zheng Chen

School of Power and Mechanical Engineering, Wuhan University,  
Wuhan, 430072 China  
xtzhuan@whu.edu.cn

**Abstract.** A modelling approach of the motion dynamics of a four-wheel electric robot is proposed. The mechanism analysis method is employed to establish a mathematical model for a robot. With experimental data of various scenarios for the step responses of the system, the parameters of the model are identified. Simulation shows the feasibility and reliability of the motion modelling approach.

**Keywords:** Wheel robot; motion modelling; parameter estimation.

## 1 Introduction

In recent years, the applications of mobile robots with four wheels, are becoming more and more popular [1]. The common controllers for them are designed based on PID algorithms and/or fuzzy logic algorithms. Without the mathematical models of the motion dynamics, the parameters in the controllers are tuned with experiments. The design process is hard and subjective to the designer's experiences. If the mathematical model is known, it is possible to design controllers with the performance optimized [2, 3].

A robot is subjected to vertical, lateral and longitude's forces, which are interactive and coupled. The integrated analysis of three degrees of freedom has appeared and the analysis is extended to nonlinear characteristics [4]. However, it is common and convenient to independently analyze the vertical, lateral and longitudinal dynamics such that the degree of freedom is decreased and it is simple. Moreover, when the robot operation condition is limited, the interactivities and coupling will be very weak.

To facilitate the control design for a four-wheel robot, the modelling of motion dynamics is studied. Both the mechanism analysis and experimental test are employed to establish a mathematical model for the studied robot. The robot is with steering wheels at the front and driving wheels at the rear. The input variables are the control command signals for the steering gear and the driving motor, which respectively manipulate the steering angle and advancing speed, while the output variables are the coordinates of the robot in the plane. Firstly,

---

<sup>\*</sup> This paper is partially supported by "the Fundamental Research Funds for the Central Universities."

the model structure is obtained through the analysis of robot mechanism. The robot model is divided into three parts: steering subsystem, driving subsystem and robot kinematics. With the dynamical functions of the three parts combined, the robot model structure is determined. Then some experimental scenarios on the robot are designed and executed to get the operation data. With data processing and parameter identification algorithms, the unknown parameters are identified. Comparing simulation with experiment results under the same settings, it proves the feasibility and reliability of the proposed approach.

The remaining part of the paper is organized as follows: Section 2 described the determination of the dynamical model structure for the robot with mechanism analysis. Section 3 presents the parameter identification of the robot model. Some conclusions are given in Section 4.

## 2 Model Structure

A four-wheeled robot, with steering wheel controlled by a DC servo at the front, and driving wheels controlled by a DC motor at the rear is studied. The rear wheels are fixed in accordance with the robot body longitudinal direction. The robot body is assumed to be rigid and the rotating dynamics are ignored.

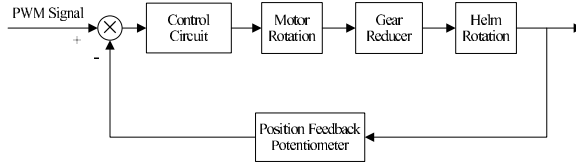
The planar motion is studied here, and the control variables are the PWM signals  $x_a(t)$ ,  $x_u(t)$  manipulating the steering gear and driving motor respectively while the controlled variables are the plane coordinates of the axis center of rear wheels  $(X_M, Y_M)$ .

The steering gear and driving motor subsystems are studied independently and then be merged together through the robot kinematical model. In the steering gear model, the input is the PWM signal for gear, and the output is the front wheel angle. In the driving motor model, the input is the PWM signal for the motor, while the output is the speed of rear wheels, which is partial affected by the front wheel angle. With the robot kinematics, the plane coordinates of the rear wheel axis center are expressed as functions of the front wheel angle and the rear wheel speed.

### 2.1 Steering Gear Model Structure

The front wheel angle is manipulated through a DC electric steering gear in the robot. This gear is controlled by a PWM signal  $x_a(t)$ . An electric steering gear is a typical position servo loop system as shown in Fig. 1.

An actual steering gear is much more complex than that shown in Fig. 1. On the other hand, a more complicated control algorithm is applied to obtain a good performance. Generally, it is difficult to know the control algorithm of an existing steering gear, and the parameters of the DC motor and gear reducer are unknown. So it is impractical to obtain the steering gear model with the analysis of the gear components. However, it is known that the steering system is a time-delay system. The closed-loop dynamics of the steering gear could be approximated by a first-order system with delay as (1).



**Fig. 1.** Electric steering servo system

$$G_{servo}(s) = \frac{K e^{-\tau s}}{Ts + 1}, \tag{1}$$

The input of the model (I) is  $x_a(t)$  while the output is the steering angle  $\varphi(t)$ . The model (II) could be expressed by the following differential equation

$$T \frac{d\varphi(t)}{dt} + \varphi(t) = f(x_a(t - \tau)). \tag{2}$$

With the geometric analysis of the front wheel steering structure, the relationship between front wheel angle  $\alpha(t)$  and steering angle  $\varphi(t)$  is as follows,

$$\alpha(t) = f_{\varphi 2\alpha}(\varphi(t)). \tag{3}$$

### 2.2 Motor Model Structure

The rear wheel of the robot is driven by a permanent magnet DC motor, whose speed is adjusted through the motor’s armature voltage. The DC drive motor circuit is an H-bridge circuit. The armature voltage is modulated by a PWM signal with different duty cycle approximating different average voltage value to achieve the speed control.

Stress analysis is required to establish the motor model. The robot resistance forces include acceleration resistance force, rolling resistance force, air resistance force and the slope resistance force. The driving force is the motor torque transmitted by the driving wheel on the transmission. The force balance is as follows,

$$F_t = \sum F, \tag{4}$$

where  $F_t$  is the driving force and  $\sum F$  is the sum of the resistance force.

**Resistance force.** When the robot is uniformly moving on a level road and the robot’s internal mechanical friction loss is ignored, the resistance primarily includes rolling resistance force  $F_f$  and air resistance force  $F_w$ . When the robot has an acceleration, then acceleration resistance force  $F_j$  emerges.

When the elastic tire is rolling on a hard road, the tire deformation is the main course for energy consumption [5]. In order to overcome the rolling dual resistance a push force is characterized as the rolling resistance forces  $F_f$  calculated as follows,

$$F_f = fmg, \tag{5}$$

where  $m$  is the robot mass,  $g$  the gravity acceleration, and  $f$  the rolling resistance force coefficient which could be calculated through the moment equilibrium condition. When the rolling speed is low,  $f$  could be seen as a constant.

Air resistance force is the decomposed part in the running direction of the air force, which could be ignored when the robot speed is low [6].

The acceleration resistance force is the product of robot mass and the acceleration speed. However, when the vehicle accelerates, not only the planar movement produces the inertial force, but also the rotating mass inertia moment produces the inertia force. In the calculation of the acceleration resistance force, a coefficient  $\delta - 1 (> 0)$  is considered to translate a rotation mass inertia moment into planar mass. So the acceleration resistance force is as follows,

$$F_j = \delta m \frac{du(t)}{dt}. \quad (6)$$

The slope resistance force is ignored in this model. With the above analysis, the sum of the resistance force undertaken by the robot is

$$\sum F = F_f + F_j = fmg + \delta \frac{du(t)}{dt}. \quad (7)$$

**Driving force.** The driving force is provided by a permanent magnetic DC motor, whose mechanical characteristic equation is

$$n = \frac{U - I_a(R_a + R_j)}{C_E \Phi} = \frac{U}{C_E \Phi} - \frac{(R_a + R_j)T_{em}}{C_E C_T \Phi^2}, \quad (8)$$

where  $n$  is the motor rotating speed, in units of  $r/min$ ,  $U$  the armature voltage,  $R_a$  the armature winding resistance,  $R_j$  the adjustable resistance series in the armature circuit,  $I_a$  the armature current,  $C_E$  the electromotive force constant,  $C_T$  the torque constant,  $T_{em}$  the electromagnetic torque, and  $\Phi$  the magnetic flux per pole. If the armature reaction effect is ignored,  $\Phi$  is constant [7].

If the shaft moment of inertia of the motor is ignored, then  $T_{em} = T_2 + T_0$ , where  $T_2$  is the mechanical torque and  $T_0$  is the non-load torque.

$$T_2 = F_t r, \quad (9)$$

where  $F_t$  is the driving force and  $r$  is the arm length of the driving wheels.

The output power of the motor is  $P_{out} = T_2 \Omega = T_2 (n2\pi)/60 = n\pi T_2/30$ .

From (8), it is seen that the speed is proportional to the armature voltage when the output torque is constant. The speed is adjustable with the change of the armature voltage  $U$ . The most common approach for speed adjustment is that the armature voltage is proportional to the duty cycle  $x_u(t)/X_{u\max}$  of a PWM signal where  $X_{u\max}$  is the control resolution. For an electric motor, when output voltage of the power supply is  $U_0$ , then

$$U = U_0 \times x_u(t)/X_{u\max}. \quad (10)$$

With (10), (8) and (9), the following is true.

$$P_{out} = \frac{F_t \pi r}{30} \left( \frac{U_0 x_u(t)}{X_{u\max} C_E \Phi} - \frac{(R_a + R_j)(T_2 + T_0)}{C_E C_T \Phi^2} \right). \quad (11)$$

**Motor model.** For a DC motor,

$$F_{out} = F_t u(t), \tag{12}$$

where  $u(t)$  the rear wheel velocity. Compared (12) with (11) and (9), the motor model could be written in the following framework,

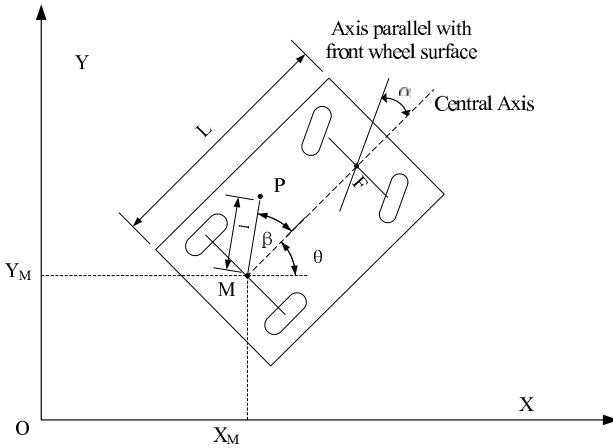
$$k_3 \frac{du(t)}{dt} + u(t) = k_1 x_u(t) - k_2, \tag{13}$$

where

$$k_1 = \frac{r\pi U_0}{30 X_{u\max} C_E \Phi}, k_2 = \frac{r\pi(R_a + R_j)}{30 C_E C_T \Phi^2} (fmg r + T_0), k_3 = \frac{(R_a + R_j)\pi r^2 \delta}{30 C_E C_T \Phi^2}.$$

### 2.3 Robot Kinematics

There are several kinds of model structures for four-wheeled robots. A two-wheeled bicycle model could be used to study turn movements and propelling dynamical control with the assumption that the steering angle and wheel side angle of the left wheels are the same as those of the right wheels. So a two-wheeled model is employed in this study.



**Fig. 2.** Robot planar movement

A planar movement of a robot is as shown in Fig. 2. Point F and M are the front wheel axis center and rear wheel axis center respectively. The variables  $L, l, \beta, \theta, \alpha$  are shown as in Fig. 2. The robot is propelled by the rear wheels, so the velocity of point M in the direction MF is the robot velocity  $u(t)$ . To simplify the model, point M is chosen as the reference and the kinematics are as follows,

$$\begin{aligned}
T \frac{d\varphi(t)}{dt} + \varphi(t) &= f(x_a(t - \tau)), & \alpha &= f_{\varphi 2\alpha}(\varphi), \\
k_3 \frac{du(t)}{dt} + u(t) &= k_1 x_u(t) - k_2, \\
\frac{dX_M}{dt} &= u(t) \cos \theta, \\
\frac{dY_M}{dt} &= u(t) \sin \theta, \\
\frac{d\theta}{dt} &= \frac{u(t)}{L} \tan \alpha.
\end{aligned} \tag{14}$$

### 3 Model Parameter Identification

From (14), there are three parameters in the steering subsystem required to be identified: the time constant  $T$ , the time delay  $\tau$  and the function  $K = f(x_a(t))$ . From the analysis of test experimental results,  $K = f(x_a(t))$  is a piecewise function. The function  $\alpha = f_{\varphi 2\alpha}(\varphi)$  could be deduced from the geometry analysis. There are three parameters in motor model required to be identified:  $k_1$ ,  $k_2$  and  $k_3$ . The parameter  $k_3$  is a time constant,  $k_1$  is only determined by the motor features and the armature voltage,  $k_2$  is constant when the road conditions are unchanging. The parameters in the kinematical equations of (14) are known from the geometry structure.

#### 3.1 Steering Gear Model Parameter Identification

The control input variable is the duty cycle  $x_a(t)$  of the PWM signal for steering servo and the output is the front wheel angle  $\alpha(t)$ . To identify the model parameters in steering system,  $\alpha(t)$  should be measured. In this subsection, a measurement method is proposed for the front wheel angle. Some experiments are done to get the running data (input and output). With those data, the parameters  $T$ ,  $\tau$  and function  $K = f_a(\alpha)$  are identified.

**Front wheel angle measurement.** There are many angle sensors, including laser angle interferometer (high-precision and expensive), rotary encoder (used in multi-turn angle measurement), non-contact angle sensor based on Hall devices and potentiometer-based rotation sensors. With the price and installation convenience considered, a direct sliding potentiometer (30mm length) is employed to calculate the angle through measuring the potentiometer displacement.

The installation is shown in Fig. 3, where grey part is the fixed shaft, straight-sliding potentiometer is fixed to the robot chassis, and its handle is laterally moving when the servo output arm rotates. by measuring the horizontal displacement of the handle from the center, the steering angle and then the wheel angle can be calculated. The sliding potentiometer has three pins, are connected to the handle terminal and the two potentiometer terminals. The potentiometer terminals are connected across the two pins on the +5V reference voltage



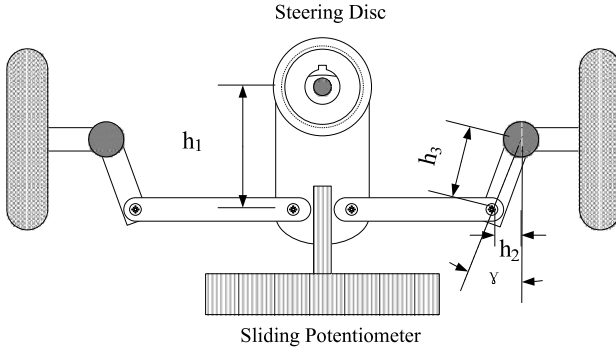


Fig. 3. Steering angle measurement  $\alpha = 0$

resource, and the third pin connected micro-controller ADC channel, by which the displacement of the handle could be read out.

From the geometry analysis in Fig. 3, when the arm of the steering system start to rotate from the vertical position, the sliding potentiometer's handle moves  $x$  to the right, the steering angle is  $\varphi = \arcsin(x/h_1)$ . When  $\varphi$  is small,  $\arcsin \varphi \approx \varphi$ , so  $x$  has approximate linear relationship with  $\varphi$ . The output voltage of the arm handle for handle displacement  $x$  is  $y_a$  (ADC output, 8-bit precision), and the total travel of the sliding potentiometer is 30 mm.

Then  $x = \frac{0.03y_a}{256}$ . From Fig. 3, it is obvious that  $\alpha = \gamma + \arcsin(\frac{x - h_2}{h_3})$ . The Parameters  $h_1$ ,  $h_2$ ,  $h_3$  and  $\gamma$  are determined by the potentiometer structure. In Fig. 3,  $\gamma = 0.253\text{rad}(14.5^\circ)$  and  $h_3 = 15.5\text{mm}$ .

In an experiment, when  $x = h_2$ ,  $y_a = 47$ . Then the following is true,

$$\alpha = 0.253 + \arcsin\left(\frac{(y_a - 47) \times 0.03}{256 \times 0.0155}\right). \quad (15)$$

With (15), the front wheel angle could be directly calculated with the ADC output, instead of the medium variable  $\varphi$ .

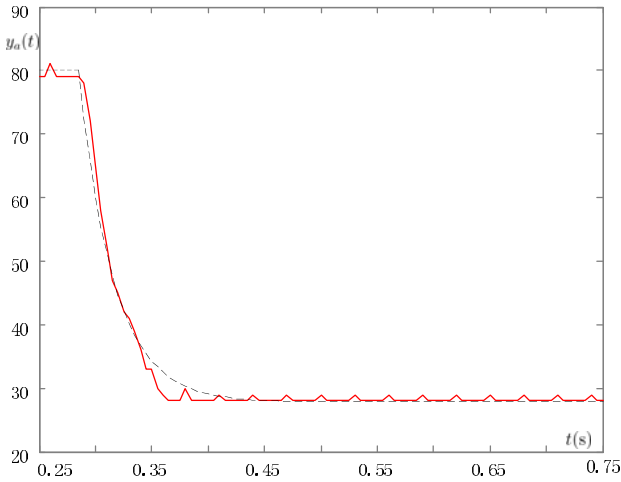
**Parameter identification.** When  $\varphi$  is small, it has a linear relationship with  $y_a$ . So the steering model could be written as follows,

$$T_a \frac{dy_a(t)}{dt} + y_a(t) = f_a(x_a(t - \tau)), \quad (16)$$

in which  $\tau$  could be read from the step response of  $y_a(t)$ .

In the experiment, when  $x_a(t) = 5600$ ,  $\varphi$  is zero, *i.e.*,  $f_a(5600) = 0$ . So in the step response experiment, the initial value of  $x_a(t)$  is equal to 5600. Furthermore, when  $x_a(t) = 5600$ ,  $y_a = 80$ . So the following is true,

$$y_a(t) = f_a(x_a(t)) + (80 - f_a(x_a(t)))e^{-\frac{t-\tau}{T_a}}. \quad (17)$$



**Fig. 4.** The output  $y_a(t)$  when  $x_a=5100, 0.25s < t < 0.75s$

The least square method is employed to estimate  $T_a$  and  $f_a(x_a)$ . Least square method is the most commonly used for parameter estimation and data fitting. The best fit in the least-squares sense minimizes the sum of squared residuals.

In the experiment, when

$$x_a(t) = \begin{cases} 0, & t < 0; \\ 5600, & 0 \leq t \leq 0.25s; \\ 5100, & 0.25s < t \leq 2.5s, \end{cases} \quad (18)$$

the output  $y_a(t)$  is shown in Fig. 4. It can be seen that when the input changes at  $t = 0.25s$ , the output has a response with a time delay. The response seems to be very quick. From Fig. 4, it can be seen that  $\tau \approx 0.035s$ . Then the least square method is applied to estimate  $T_a$  and  $f_a(x_a)$ .

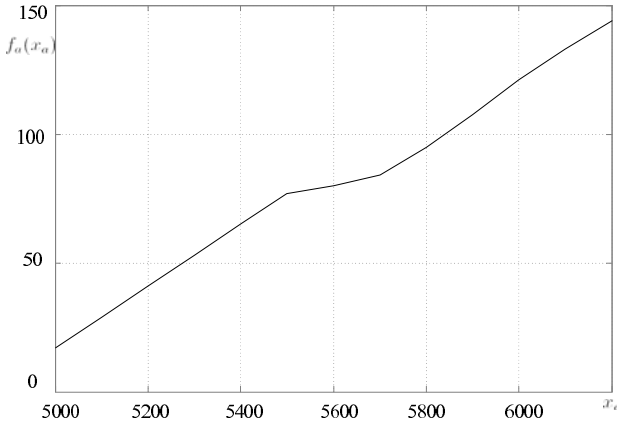
The estimation result is that  $T_a = 0.0311s$  and when  $x_a = 5100$   $y_a(\infty) = f_a(x_a) = 27.8125$ . So the solution of (16) is  $y_a(t) = 27.8125 + (80 - 27.8125)e^{\frac{t-0.035}{0.0311}}$ .

The simulation result is shown as the dash line in Fig. 4.

More experiments are done with 5100 in (18) replaced by 5000, 5100, 5200, 5300, 5400, 5800, 5900, 6000, 6100, 6200. The average time delay is 0.03s, *i.e.*,  $\tau = 0.03s$ . The average time constant is 0.03s, so  $T_a = 0.03s$ . The values of  $f_a(x_a)$  is plotted in Fig. 5 with the horizontal axis is  $x_a$ .

The piecewise polynomial fitting with least square method is as follows,

$$f_a(x_a) = \begin{cases} 0.12x_a - 583, & x_a \leq 5500, \\ 0.03x_a - 88, & 5500 < x_a \leq 5700, \\ 0.122x_a - 612.7, & x_a > 5700. \end{cases} \quad (19)$$



**Fig. 5.** Function of  $f_a(x_a)$  with respect to  $x_a$

The servo model (16) is as follows,

$$0.03 \frac{dy_a(t)}{dt} + y_a(t) = f_a(x_a(t - 0.03)). \quad (20)$$

Then the steering subsystem model is the combination of (15), (19) and (20).

### 3.2 Motor Model Parameter Identification

The input variable is  $x_u(t)$ , the duty cycle of the PWM signal for the armature voltage while the output variable is the robot velocity. The parameters required to be identified in (12) are  $k_1$ ,  $k_2$ ,  $k_3$ .

**Robot velocity measurement.** There are many methods for velocity measurement. The principle is to transfer the wheel speed into a pulse signal, and then count the pulse number within a unit time such that the velocity could be calculated. A rotary encoder speed sensor is employed in the following experiments.

In the experiments, the rotary encoder output pulse number is  $\frac{100 \times 76}{18}$  once the rear wheel rotates  $360^\circ$ . The semidiameter of the wheel is 0.026m. If the pulse count is  $y_u$  within 0.01s, then the robot velocity is as follows,

$$u(t) = 2\pi \times 0.026 \times \frac{y_u}{100} \times \frac{18}{76} \times \frac{1}{0.01} = 0.0387y_u. \quad (21)$$

**Parameter identification.** With  $k_{y1} = \frac{k_1}{0.0387}$ ,  $k_{y2} = \frac{k_2}{0.0387}$  and  $k_{y3} = k_3$ , (13) could be written as

$$y_u(t) + k_{y3} \frac{dy_u(t)}{dt} = k_{y1}x_u(t) - k_{y2}. \quad (22)$$

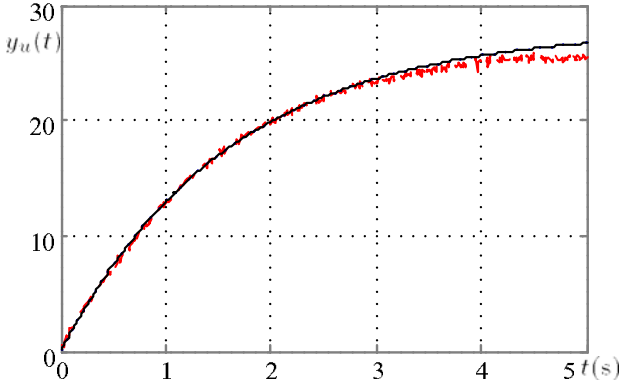


Fig. 6. Motor step response when  $x_u(t) = 40$

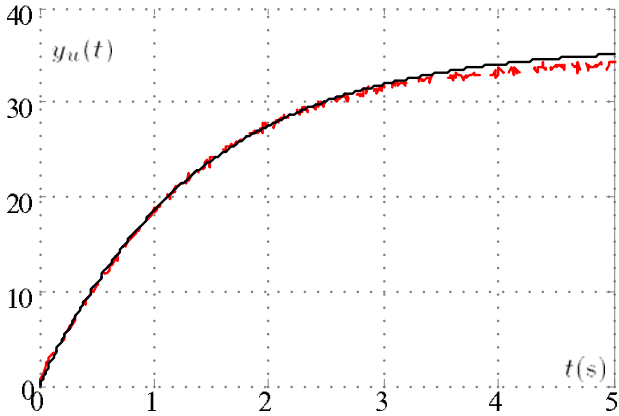


Fig. 7. Motor step response when  $x_u(t) = 50$

With the disturbance considered, a set of experiment will be done and then the average sampled value will be used for parameter identification for every input series of  $x_u(t)$ . For example, there is a set of experiments (10 groups) with the input setting  $x_u(t) = 40, x_a(t) = 5600(\alpha = 0)$ . The average value  $\bar{y}_u(t)$  of 10 groups output are applied for parameter estimation. The estimation result is that  $k_{y3} = 1.61s$ , and  $k_{y1}x_u - k_{y2} = 27.93$ , which indicates the motor model is

$$y_u = 27.93(1 - e^{-\frac{t}{1.61}}).$$

The experiment result and simulation are shown respectively as the solid line and dash line in Fig. 6.

With different settings, respective experiments are done as follows. When  $x_u = 50$ ,  $k_{y3} = 1.39s$ , and  $k_{y1}x_u - k_{y2} = 36.05$ .

When  $x_u = 60$ ,  $k_{y3} = 1.46\text{s}$ , and  $k_{y1}x_u - k_{y2} = 47.77$ .

When  $x_u = 70$ ,  $k_{y3} = 1.24\text{s}$ , and  $k_{y1}x_u - k_{y2} = 56.61$ .

The experiment result and simulation are shown respectively as the solid line and dash line in Figs. 7, 8 and 9. Then the average value of  $k_{y3}$  is  $\bar{k}_{y3} = 1.43\text{s}$ . The variable  $k_{y1}x_u - k_{y2}$  is plotted in Fig. 10. With least square method employed for polynomial fitting,  $k_{y1} = 0.98$  and  $k_{y2} = 11.68$ .

Then with the front wheel angle  $\alpha = 0$ , the motor driving model is

$$\begin{aligned}
 1.43 \frac{dy_u(t)}{dt} + y_u(t) &= 0.98x_u(t) - 11.68, \\
 u(t) &= 0.0387y_u(t).
 \end{aligned}
 \tag{23}$$

It is noticed that the above model is only applicable when  $u > 0$ .

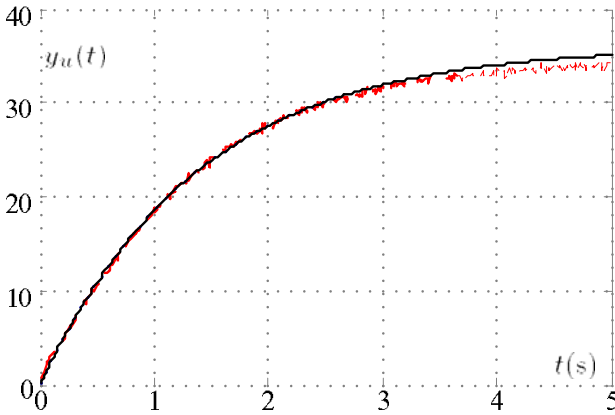


Fig. 8. Motor step response when  $x_u(t) = 60$

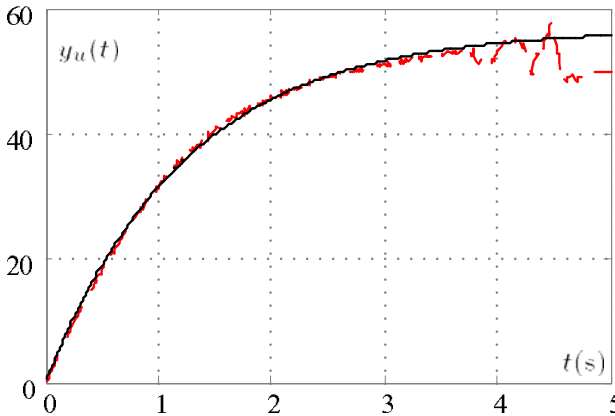
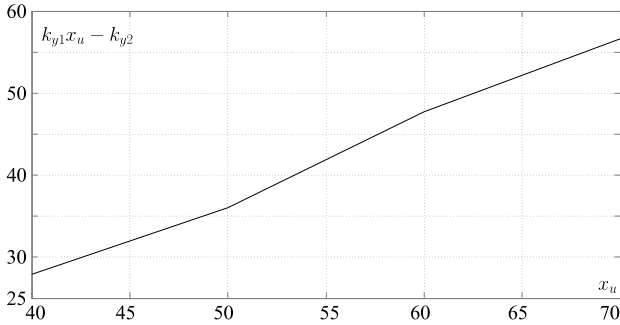


Fig. 9. Motor step response when  $x_u(t) = 70$



**Fig. 10.**  $k_{y1}x_u - k_{y2}$  vs.  $x_u$

Some simulation and experiments have been done for model validation. By comparing model simulation results and the actual experiments, the robot model is proved to be feasible and reliable.

## 4 Conclusion

The motion dynamics of a four-wheel electric robot is modelled. The model structure is determined with mechanism analysis while the parameters in the model are identified from the experimental data with least squares method applied. The input signals of the model are the steering gear control signal and motor control signal, and the output of this model is the complanate coordinate of the robot. With the simulation (using the identified model) and the actual experimental data (using the robot running with the same setting) compared, it can be seen the feasibility and reliability of the proposed modeling approach.

## References

- [1] Ichihara, Y., Ohnishi, K.: Path planning and tracking control of wheeled mobile robot considering robot's capacity. In: Proceedings of the IEEE International Conference on Industrial Technology, Mumbai, India, pp. 176–181 (2006)
- [2] Dincmen, E., Acarman, T.: Application of vehicle dynamics' active control to a realistic vehicle model. In: American Control Conference, New York, pp. 200–205 (2007)
- [3] You, F., Wang, R., Zhang, R.: Study of system identification and control algorithm for intelligent vehicle. China Journal of Highway and Transport 21(4), 111–116 (2008)
- [4] Silva, L., Magallan, G., Angelo, C.D., Garcia, G.: Vehicle dynamics using multi-bond graphs: Four wheel electric vehicle modeling. In: IECON 2008 - 34th Annual Conference of IEEE Industrial Electronics, Orlando, Florida, pp. 2846–2851 (2008)
- [5] Yu, Z.: Automobile theory, pp. 2–21. Machinery Industry Press, Beijing (2006)
- [6] Wang, F.: The mathematics model of the motion velocity of army vehicle. Journal of Institute of Surveying and Mapping (1), 66–70 (1993)
- [7] Gu, C., Chen, Q., Xiong, Y.: Electrical Machines, 2nd edn., pp. 82–94. Huazhong University of Science and Technology Press, Wuhan (2005)

# Sliding Mode Control of Hybrid Joints for Wheeled Mobile Manipulators with Markovian Switching and Bounded Torques\*

Lei Dai, Dan Huang, and Zhijun Li\*\*

Department of Automation, Shanghai Jiao Tong University, Shanghai, China, 200240  
zjli@ieee.org

**Abstract.** The hybrid joints proposed in [1] and [2] can be switched to either active (actuated) or passive (under-actuated) mode as needed. In this paper, we develop a sliding mode control via Markovian switching for wheeled mobile manipulators with hybrid joints. The switching rate uncertainties are allowed within an uncertainty domain. By introducing specified matrices, the connections among the designed sliding surfaces corresponding to each mode are established. Furthermore, the presented sliding mode controller including the transition rates of modes can cope with the effect of Markovian switching, and the bounded torque control is designed to follow the desired trajectory for wheeled mobile manipulators subject to joint velocity constraints. Finally, a numerical example is given to illustrate the applicability of the method.

## 1 Introduction

The hybrid joint shown in Fig. 1 was first proposed in [1] and [2], which is with one clutch and one brake. When the clutch is released, the link is free, and the passive link is directly controlled by the dynamic coupling of mobile manipulators; when it is on, the joint is actuated by the motor. Moreover, the passive link can be locked by the brake as needed. The robot with hybrid joints is called as the hybrid actuated robot.

One of the advantages of using hybrid actuated robots is that they may consume less energy than the fully-actuated ones. For example, hyper-redundant robots, such as snake-like robots or multi-legged mobile robots, need large redundancy for dexterity and specific task completion; under-actuation allows a more compact design, much simpler control and communication schemes. The hybrid actuated robot concept is also useful for the reliability or fault-tolerant design of fully-actuated robots working in hazardous areas or with dangerous materials. If any of the joint actuators of such a device fails, one degree of freedom of the system would be lost. It is important, in these cases, that the passive (failed) joint can still be controlled via the dynamic coupling with the active

---

\* This work is supported by Shanghai Pujiang Program under Grant No. 08PJ1407000, Natural Science Foundation of China under Grant Nos. 60804003 and 60935001.

\*\* Corresponding author.

ones, so the system can still make use of all of its degrees of freedom originally planned.

Hybrid actuated mobile manipulator [9], [10] is the robot manipulator consisting of hybrid joints mounting on a wheeled mobile robot, which first appeared in [1] and [2]. Hybrid actuated mobile manipulators are different from full-actuated mobile manipulators in [5], due to simultaneously integrating both kinematic constraints and dynamic constraints. The hybrid joint is also with the characteristic of under-actuated the joints [4], [3], for example, the hybrid joints in the free mode, which can rotate freely, can be indirectly driven by the effect of the dynamic coupling between the active and passive joints. The zero torque at the hybrid joints results in a second-order nonholonomic constraint. The effective control of second-order nonholonomic system could reduce the weight, energy consumption and cost.

However, hybrid joint switching [9], [10] might occur suddenly dependent on the task requirement during the motion. To our best knowledge, there is no attention paid to the hybrid joint switching of wheeled mobile manipulators, because the mobile manipulator incorporating the mobile robot and robotic manipulators, the joint switching seems to be stochastic and the switching may appear in any joints of the robot, while simple switching approach can not handle all the possibility. To avoid the necessity of stopping the robot as a joint switches, the Markov theory was developed to design a procedure to incorporate abrupt changes in the hybrid joints of mobile manipulators, and the Markovian jumping systems theory shall be used to guarantee the stochastic stability. The Markovian jump linear system (MJLS) [6] and [7], is an important class of stochastic dynamic systems that is popular in modeling abrupt changes in the system structure. It is a hybrid system which is loosely defined as a system that involves the interaction of both discrete events (represented by finite automata) and continuous-time dynamics (represented by differential equations). By contrast, the problem of sliding mode control (SMC) for stochastic systems with Markovian switching has received relatively little attention. This paper will be concerned with the design of SMC for mobile manipulators using hybrid joints with Markovian switching. There exist parameter uncertainties, nonlinearities, and external disturbance in the systems under consideration.

## 2 System Description

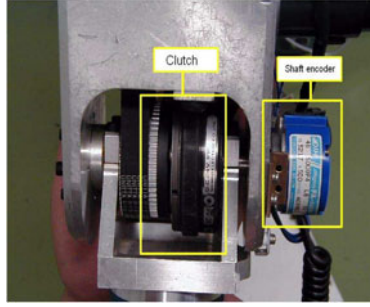
### 2.1 Dynamics

Consider an  $n_a$  DOF robotic manipulator mounted on a two-wheeled driven mobile platform, the dynamics can be described as

$$M(q)\ddot{q} + V(q, \dot{q})\dot{q} + G(q) + d(t) = B(q)\tau + f \quad (1)$$

where  $q = [q_v^T, q_a^T]^T \in \mathbb{R}^n$  with  $q_v = [x, y, \vartheta]^T \in \mathbb{R}^{n_v}$  denoting the generalized coordinates for the mobile platform and  $q_a \in \mathbb{R}^{n_a}$  denoting the coordinates of the robotic manipulator joints, and  $n = n_v + n_a$ . Eq. (1) includes the symmetric positive definite inertia matrix  $M(q) \in \mathbb{R}^{n \times n}$ , the Centripetal and Coriolis





**Fig. 1.** The hybrid joint

torques  $V(q, \dot{q}) \in \mathbb{R}^{n \times n}$ , the gravitational torque vector  $G(q) \in \mathbb{R}^n$ , the known input transformation matrix  $B(q) \in \mathbb{R}^{n \times m}$ , the control inputs  $\tau \in \mathbb{R}^m$  and the generalized constraint forces  $f \in \mathbb{R}^n$ , and  $J_v \in \mathbb{R}^{l \times n_v}$  is the kinematic constraint matrix related to nonholonomic constraints;  $\lambda_n \in \mathbb{R}^l$  is the associated Lagrangian multipliers with the generalized nonholonomic constraints. We assume that the mobile manipulator is subject to known nonholonomic constraints. In actual implementation, we can adopt the methods of producing enough friction between the wheels of the mobile platform and the ground.

## 2.2 Joint Velocity Constraints

Each and every individual joint velocity is bounded as follows

$$|\dot{q}_i(t)| \leq B_i, t \geq 0, i = 1, \dots, n \quad (2)$$

where  $B_i > 0$  is a constant. Choosing  $\Omega_i = B_i^2/2$ , the above velocity constraints can be expressed as

$$\frac{1}{2} \dot{q}_i^2(t) \leq \Omega_i, t \geq 0, i = 1, \dots, n \quad (3)$$

## 2.3 Reduced System

When the system is subjected to nonholonomic constraints, the  $(n - m)$  nonintegrable and independent velocity constraints can be expressed as

$$J_v(q) \dot{q}_v = 0 \quad (4)$$

The constraint (4) is referred to as the classical nonholonomic constraint when it is not integrable. In the paper, constraint (4) is assumed to be completely nonholonomic and exactly known.

Since  $J_v(q) \in \mathbb{R}^{(n_v - m) \times n}$ , introduce  $J_a \in \mathbb{R}^{n_a \times n}$ , and  $J = [J_v, J_a]^T \in \mathbb{R}^{(n - m) \times n}$ , such that it is possible to find a  $m + n_a$  rank matrix  $R(q) \in \mathbb{R}^{n \times (m + n_a)}$  formed by a set of smooth and linearly independent vector fields spanning the null space of  $J(q)$ , i.e.,  $R^T(q) J^T(q) = 0$ , such that

$$\dot{q} = R(q)\dot{z}(t) \tag{5}$$

Differentiating (5) yields

$$\ddot{q} = \dot{R}(q)\dot{z} + R(q)\ddot{z} \tag{6}$$

From (5),  $\dot{z}$  can be obtained from  $q$  and  $\dot{q}$  as  $\dot{z} = [R^T(q)R(q)]^{-1}R^T(q)\dot{q}$ . The dynamic equation (1), which satisfies the nonholonomic constraint (4), can be rewritten in terms of the internal state variable  $\dot{z}$  as

$$M(q)R(q)\ddot{z} + V^*\dot{z} + G(q) + d(t) = B(q)\tau + J^T(q)\lambda \tag{7}$$

with  $V^* = [M(q)\dot{R}(q) + V(q, \dot{q})R(q)]$ .

Substituting (5) and (6) into (1), and then pre-multiplying (1) by  $R^T(q)$ , the constraint matrix  $J^T(q)\lambda$  can be eliminated. As a consequence, we have the transformed nonholonomic system

$$\mathcal{M}(q)\ddot{z} + \mathcal{V}(q, \dot{q})\dot{z} + \mathcal{G}(q) + \mathcal{D} = \mathcal{U} \tag{8}$$

where  $\mathcal{M}(q) = R^T M(q)R$ ,  $\mathcal{V}(q, \dot{q}) = R^T [M(q)\dot{R} + V(q, \dot{q})R]$ ,  $\mathcal{G}(q) = R^T G(q)$ ,  $u = B(q)\tau$ ,  $\mathcal{D} = R^T d(t)$ ,  $\mathcal{U} = R^T B(q)\tau$ , which is more appropriate for the controller design as the constraint  $\lambda$  has been eliminated from the dynamics.

*Remark 1.* In this paper, we choose  $z = [\theta_r \ \theta_l \ \theta_1 \ \theta_2 \ \cdots \ \theta_{n_a}]^T$ , and  $\eta = [\theta_r, \theta_l]^T$ , where  $\theta_r, \theta_l$ : the rotation angle of the left wheel and the right wheel of the mobile platform, respectively, and  $\theta_1, \cdots, \theta_{n_a}$  are the joint angles of the link 1, 2,  $\cdots$ ,  $n_a$ , respectively, and  $\mathcal{U} = [\tau_r, \tau_l, \tau_1, \cdots, \tau_{n_a}]$ .

*Remark 2.* The total degree of freedom for a two-wheeled driven mobile manipulator with two wheels and  $n_a$  joints is  $n_q = n_a + 2$ .

**Assumption 2.1.** *The hybrid joints are only within each actuator of the mobile platform and the manipulator, such that switching may appear in every joint independently.*

*Remark 3.* The hybrid joint is equipped with each actuator of the wheels and links, since the left wheel and right wheel is symmetric, for simplification, we assume switching only happens in the left wheel and each joint of the manipulator independently.

*Remark 4.* Under Assumption 2.1, there are  $2^{n_a+1}$  modes of operation, which is listed in Table 1 depending on which hybrid joint is in the active (actuated) or passive (under-actuated) mode.

From [3], it is obvious that no more than  $h_a$  actuated hybrid joint of the mobile manipulator can be controlled at every instant. Let  $h_p$  be the number of passive hybrid joints that have not already reached their set point in a given instant. If  $h_p > h_a$ ,  $h_a$  passive joints are controlled and grouped in the vector  $z_p \in \mathbb{R}^{h_a}$ ,

**Table 1.** The modes of hybrid joints

Mode	Hybrid Joints					
	Left Wheel	Right Wheel	Joint 1	Joint 2	...	Joint $n_m$
1	actuated	actuated	actuated	actuated	...	actuated
2	actuated	actuated	passive	actuated	...	actuated
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2^{n_a} - 1$	actuated	actuated	passive	passive	...	passive
$2^{n_a}$	actuated	passive	actuated	actuated	...	actuated
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2^{n_a+1}$	actuated	passive	passive	passive	...	passive

the remaining passive hybrid joints, if any, are kept locked by the clutches, whose function is like a brake, and the active joints are grouped in the vector  $z_a \in \mathbb{R}^{h_a}$ . If  $h_p < h_a$ , the  $h_p$  passive hybrid joints are controlled applying torques in  $h_a$  active hybrid joints. In this case,  $z_p \in \mathbb{R}^{h_p}$  and  $z_a \in \mathbb{R}^{h_a}$ . The strategy is to control all passive hybrid joints until they reach the desired position, considering the conditions exposed above, and then turn on the clutch. After that, all the active hybrid joints are controlled by themselves as a fully actuated robot.

We partition the dynamics (8) into two parts, the actuated part and the passive part, represented by “a” and “p”, respectively. Then we can rewrite the dynamics (8) as

$$\begin{bmatrix} \mathcal{M}_a(\zeta) & \mathcal{M}_{ap}(\zeta) \\ \mathcal{M}_{pa}(\zeta) & \mathcal{M}_p(\zeta) \end{bmatrix} \begin{bmatrix} \ddot{z}_a \\ \ddot{z}_p \end{bmatrix} + \begin{bmatrix} \mathcal{V}_a & \mathcal{V}_{ap} \\ \mathcal{V}_{pa} & \mathcal{V}_p \end{bmatrix} \begin{bmatrix} \dot{z}_a \\ \dot{z}_p \end{bmatrix} + \begin{bmatrix} \mathcal{G}_a \\ \mathcal{G}_p \end{bmatrix} + \begin{bmatrix} \mathcal{D}_a(t) \\ \mathcal{D}_p(t) \end{bmatrix} = \begin{bmatrix} \mathcal{U}_a \\ 0 \end{bmatrix} \quad (9)$$

where  $\mathcal{M}_a \in \mathbb{R}^{h_a \times h_a}$ ,  $\mathcal{M}_{ap} \in \mathbb{R}^{h_a \times h_p}$ ,  $\mathcal{M}_{pa} \in \mathbb{R}^{h_p \times h_a}$ ,  $\mathcal{M}_p \in \mathbb{R}^{h_p \times h_p}$ : the coupling inertia matrices of the operational parts and the failed parts;  $\mathcal{V}_a \in \mathbb{R}^{h_a \times h_a}$ ,  $\mathcal{V}_{ap} \in \mathbb{R}^{h_a \times h_p}$ ,  $\mathcal{V}_{pa} \in \mathbb{R}^{h_p \times h_a}$ ,  $\mathcal{V}_p \in \mathbb{R}^{h_p \times h_p}$ : the Centripetal and Coriolis torque matrices of the operational parts and the failed parts;  $\mathcal{G}_a \in \mathbb{R}^{h_a}$ ,  $\mathcal{G}_p \in \mathbb{R}^{h_p}$ : the gravitational torque vector for the operational parts and the failed parts respectively;  $\mathcal{D}_a(t) \in \mathbb{R}^{h_a}$ ,  $\mathcal{D}_p(t) \in \mathbb{R}^{h_p}$ : the bounded external disturbance from the environments on the operational parts and the failed parts respectively;  $\tau_a \in \mathbb{R}^{h_a}$ : the control input torque vector for the operational parts of the manipulator;  $\mathcal{U}_a \in \mathbb{R}^{h_p}$ : the control input torque vector for the failed parts of the manipulator satisfying  $\mathcal{U}_p = 0$ .

After some simple manipulation, we can obtain

$$\mathcal{U} = \bar{M}(z)\ddot{z}_p + \bar{H}(z, \dot{z}) + \bar{D}(t) \quad (10)$$

where  $\bar{M} = \mathcal{M}_a - \mathcal{M}_a \mathcal{M}_{pa}^{-1} \mathcal{M}_p$ ,  $\bar{H} = \bar{V}_1 \dot{z}_a + \bar{V}_2 \dot{z}_p + G_a - \mathcal{M}_a \mathcal{M}_{pa}^{-1} G_p$ ,  $\bar{D}(t) = D_a - \mathcal{M}_a \mathcal{M}_{pa}^{-1} D_p$ ,  $\bar{V}_1 = V_a - \mathcal{M}_a \mathcal{M}_{pa}^{-1} V_{pa}$ ,  $\bar{V}_2 = V_{ap} - \mathcal{M}_a \mathcal{M}_{pa}^{-1} V_p$ .

Before the design of sliding mode control incorporating Markov Jump Theory, linear mathematical models must first be obtained. The linearization may be done by expanding the nonlinear function into a Taylor series about the operating point and neglecting the higher order terms of the expansion. For this case, the

linear equations of motion are obtained by linearization of the general expression given by (10) about a time-varying reference trajectory or an equilibrium point, for instance, in the operation point  $z_o$  and  $\dot{z}_o$  in the reference trajectory, the fully operational mobile manipulator can be represented by (10). Let  $\bar{U} = \mathcal{U}_a - \bar{D}(t)$ , and linearizing the dynamics (10) around an operation point  $(z_o, \dot{z}_o)$  by Taylor series expansion, we obtain the following linearized system

$$\dot{x} = \bar{A}(z_o, \dot{z}_o)x + \bar{B}(z_o)\mathcal{U} \quad (11)$$

where

$$\bar{A}(z_o, \dot{z}_o) = \left[ \begin{array}{c} 0 \\ -\frac{\partial}{\partial z^T}(\bar{M}^{-1}\bar{H}(z, \dot{z})) - \bar{M}^{-1}(z)\frac{\partial}{\partial \dot{z}^T}(\bar{H}(z, \dot{z})) \end{array} \right] \Bigg|_{z_o, \dot{z}_o}$$

$$\bar{B}(z_o) = \left[ \begin{array}{c} 0 \\ \bar{M}^{-1}(z) \end{array} \right] \Bigg|_{z_o}$$

where  $x = [z_p^d - z_p, \dot{z}_p^d - \dot{z}_p]^T$  represents the state tracking error vector.

**Assumption 2.2.** *The desired trajectories  $z_p^d$  and their time derivatives up to the 3rd order are continuously differentiable and bounded for all  $t \geq 0$ .*

*Remark 5.* Consider (2) and Assumption 2.2, we have  $\|x\| \leq B_x$  with constant  $B_x$ .

### 3 Control Design

From the model described above, we investigate the following Markovian jump system

$$\dot{x}(t) = (\bar{A}(r_t) + \Delta\bar{A}(r_t))x(t) + (\bar{B}(r_t) + \Delta\bar{B}(r_t))\bar{U}(t) \quad (12)$$

$$\bar{U} = \mathcal{U}_a - \bar{D}(t) \quad (13)$$

where  $x(t)$  is the state tracking error vector,  $\bar{U}$  represents the control input.  $\bar{A}(r_t)$  is the system characteristic matrix while  $\bar{B}(r_t)$  is the full rank input matrix. Further,  $\Delta\bar{A}(r_t)$  and  $\Delta\bar{B}(r_t)$  represent the uncertainties of the system matrix and input matrix respectively. When  $r(k) = i$ ,  $i$  is one of the Markovian jumping parameters in the limited set  $S = \{1, 2, \dots, N\}$  with the mode transition rate matrix  $\Pi = (\pi_{ij})$ ,  $(i, j \in N)$ . The jump transition probability can be defined as

$$P(r_{t+\Delta t} = j | r_t = i) = \begin{cases} \pi_{ij}\Delta t + o(\Delta t), & i \neq j \\ 1 + \pi_{ii}\Delta t + o(\Delta t), & i = j \end{cases} \quad (14)$$

where  $\sum_{i=1, j \neq i}^N \pi_{ij} = -\pi_{ii}$ ,  $\pi_{ij} \geq 0, \forall i, j \in \Omega, i \neq j$ . Here,  $\Delta t > 0$  and  $\lim_{\Delta t \rightarrow 0} o(\Delta t)/\Delta t = 0$ .

**Assumption 3.1.** *The matrix couple  $(A_i, B_i)$  is stabilizable.*

**Assumption 3.2.** *The input matrix  $B_i$  has full rank and satisfies  $m < n$ .*

**Assumption 3.3.** *There are known constants  $\rho_{A_i}$ ,  $\rho_{B_i}$  and constant function  $\rho_f(x, t)$  enabling  $\|\Delta A_i\| \leq \rho_{A_i}$ ,  $\|\Delta B_i\| \leq \rho_{B_i}$  while  $\|\bar{D}_t\| \leq \rho_f(x, t)$  holds for all  $t$ .*

**Lemma 1.** [8] *For any vectors  $a$  and  $b$  with appropriate dimensions and any  $W > 0$ , there are  $2\sqrt{\|a\|}\sqrt{\|b\|} \leq \|a\| + \|b\|$ ,  $2a^T b \leq a^T W a + b^T W^{-1} b$ .*

**Lemma 2.** [8] *For the matrix  $A$  and  $B$  with appropriate dimensions, if  $(I + AB)$  is nonsingular, then  $(I + AB)^{-1} = I - A(I + BA)^{-1}B$ .*

**Theorem 1.** *Given a Markov jump linear system with the system parameter matrices  $A_i, B_i, C_i, D_i$ , and  $I > \eta^2 D_i^T D_i$ , for  $\eta \geq 0$ ,  $\Phi(t)$  is unknown but satisfying  $\|\Phi(t)\| \leq \eta$ ,*

$$\dot{x} = [A_i + B_i[I - \Phi(t)D_i]^{-1}\Phi(t)C_i]x \quad (15)$$

if there exists  $P_i > 0$  satisfies the following inequality for each  $i \in S = 1, 2, \dots, N$

$$\begin{bmatrix} P_i A_i + A_i^T P_i + \sum_{j=1}^N \pi_{ij} P_j & \eta P_i B_i & C_i^T \\ \eta B_i^T P_i & -I & \eta D_i^T \\ C_i & \eta D_i & -I \end{bmatrix} < 0 \quad (16)$$

then the system (15) is stable in the mean square sense.

*Proof.* The proof is similar as Lemma 3 in [8], we omit it here.

Define the transfer matrix  $M_i$  and the related vector  $v$ , we have

$$M_i = \begin{bmatrix} (\tilde{B} X_i \tilde{B}) \tilde{B}^T \\ (\tilde{B}^T X_i^{-1} \tilde{B})^{-1} \tilde{B}^T X_i^{-1} \end{bmatrix} = \begin{bmatrix} V_i \\ S_i \end{bmatrix} \quad (17)$$

$$v = [v_1 \ v_2]^T = M_i x \quad (18)$$

where  $v_1 \in R^{n-m}$ , and  $v_2 \in R^m$ ,  $\tilde{B}$  is any basis of the null space of  $B^T$ , i. e.,  $\tilde{B}$  is an orthogonal complement of  $B$ , Note that given any  $B$ ,  $\tilde{B}$  is not unique. Moreover,  $M_i^{-1} = [X_i \tilde{B}_i \ \tilde{B}_i]^T$ .

With the above transfer matrix, we can obtain the sliding mode system as follows [8]:

$$\begin{bmatrix} \dot{v}_1 \\ \dot{v}_2 \end{bmatrix} = \begin{bmatrix} \bar{A}_{i1} & \bar{A}_{i2} \\ 0 & -\gamma I \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad (19)$$

with  $\gamma \geq 0$ , and

$$\bar{A}_{i1} = (\tilde{B}^T X_i \tilde{B})^{-1} \tilde{B}_i^T (I - \Delta \tilde{B}_i (I + S_i \Delta \tilde{B}_i)^{-1} S_i) (\bar{A}_i + \Delta \bar{A}_i) X_i \tilde{B}_i$$

$$\bar{A}_{i2} = (\tilde{B}_i^T X_i \tilde{B}_i^T)^{-1} (I - \Delta \tilde{B}_i (I + S_i \Delta \tilde{B}_i)^{-1} S_i) (\bar{A}_i + \Delta \bar{A}_i) \tilde{B}_i + \gamma (\tilde{B}_i^T X_i \tilde{B}_i^T)^{-1} \tilde{B}_i^T (I + S_i \Delta \tilde{B}_i)^{-1}$$

**Theorem 2.** *If the following linear matrix inequality have the solution  $(X_i, d_{i0}, d_{i1}I, d_{i2}, \delta_i, \eta)$  for given  $A_i, B_i, \rho_{iA}, \rho_{iB}$ ,*

$$\begin{bmatrix} X_i & * & 0 & 0 \\ I & d_{i1} & 0 & 0 \\ 0 & 0 & d_{i2}I - X_i & 0 \\ 0 & 0 & 0 & 2\eta\lambda_B - \rho_B(d_{i1} + d_{i2}) \end{bmatrix} > 0$$

$$\begin{bmatrix} \tilde{B}_i^T(A_i X_i + X_i A_i^T + \sum_{j=1}^N \pi_{ij} X_j + d_{i0}I)\tilde{B}_i & * & * & * & * \\ \eta\tilde{B} & -I & \eta I & 0 & 0 \\ A_i X_i \tilde{B}_i & \eta I & -(1 - \delta_i)I & 0 & 0 \\ \rho_{iA} X_i \tilde{B}_i & 0 & 0 & -d_{i0}I & 0 \\ \rho_{iA} X_i \tilde{B}_i & 0 & 0 & 0 & -\delta_i I \end{bmatrix} < 0$$

where  $\lambda_{iB} = \sqrt{\lambda_{\min}(B_i^T B_i)}$ , under the sliding mode control

$$U_a = -S_i \bar{A}_i x - \gamma S_i x - \rho_t \text{Tanh}(\sigma_i) \quad (20)$$

where  $S_i = (\bar{B}_i^T X_i^{-1} \bar{B}_i)^{-1} \bar{B}_i^T X_i^{-1}$ ,  $\rho_t = \frac{1}{(1-\eta_{i0})\text{Tanh}(\sigma_i)}(\rho_A \|S_i\| \|x\| + (1 + \eta_0)\rho_f(x, t) + \eta_0 \|S_i(\bar{A}_i + \gamma I)x\| + \epsilon)$ ,  $\eta_{i0} = \rho_B \|S_i\|$ ,  $\sigma_i = S_i x + y = (\bar{B}_i^T X_i^{-1} \bar{B}_i)^{-1} \bar{B}_i^T X_i^{-1} x + y$ , and  $\gamma > 0$ ,  $\epsilon > 0$ ,  $\text{Tanh}(\sigma_i) = [\text{tanh}(\sigma_{i1}), \dots, \text{tanh}(\sigma_{in})]^T$ , while  $y$  satisfies  $\dot{y} = \gamma S_i x$ ,  $y(0) = -S_i x(0)$ , the sliding mode system (19) is asymptotically stable in mean square sense.

*Proof.* The proof is similar as Theorem 1 in [8], we omit it here.

**Theorem 3.** *Under the control law (20), the system sliding mode of the system (12) can reach the sliding mode surface and maintain on the surface.*

*Proof.* Define the following Lyapunov function

$$V = \sum_{j=1}^n \ln \cosh(\sigma_{ij}) \quad (21)$$

Taking  $\mathcal{L}$  as the infinite-simal generator, according to (21), then we obtain  $\mathcal{L}V$

$$\begin{aligned} \mathcal{L}V &= \text{Tanh}(\sigma_i) \dot{\sigma}_i \\ &= \text{Tanh}(\sigma_i) S_i [\bar{A}_i + \Delta \bar{A}_i] x + \text{Tanh}(\sigma_i) [I + S_i \Delta \bar{B}_i] (U_a - \mathcal{D}(t)) + \gamma \text{Tanh}(\sigma_i) S_i x \end{aligned}$$

Substituting the control law into the above equation, we can obtain

$$\begin{aligned} \mathcal{L}V &= \text{Tanh}(\sigma_i) S_i [\bar{A}_i + \Delta \bar{A}_i] x - \text{Tanh}(\sigma_i) [I + S_i \Delta \bar{B}_i] \mathcal{D}(t) + \gamma \text{Tanh}(\sigma_i) S_i x \\ &\quad + \text{Tanh}(\sigma_i) U_a + \text{Tanh}(\sigma_i) S_i \Delta \bar{B}_i U_a \\ &\leq -\rho_t \|\text{Tanh}(\sigma_i)\|^2 + \|\text{Tanh}(\sigma_i)\| \rho_A \|S_i\| \|x\| + \|\text{Tanh}(\sigma_i)\| (1 + \eta_{i0}) \rho_f(x, t) \\ &\quad + \|\text{Tanh}(\sigma_i)\| \eta_{i0} \|S_i(\bar{A}_i + \gamma I)\| \|x\| + \|\text{Tanh}(\sigma_i)\|^2 \eta_{i0} \rho_t \\ &\leq -\epsilon \|\text{Tanh}(\sigma_i)\|^2 \end{aligned} \quad (22)$$

Then, we can find under the proposed sliding mode control, the system state can reach the sliding mode surface in finite time and maintain on the sliding mode surface.

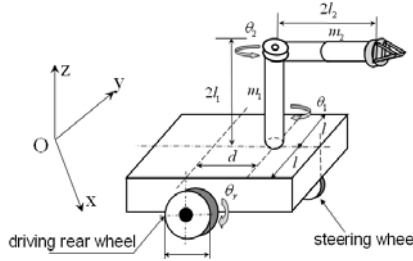
*Remark 6.* If reducing the control requirement, we can utilize the following switch law to take the place of the control (20).

$$U_a = -\frac{1}{(1-\eta_{i0})}((\rho_A \|S_i\| + \|S(\bar{A}_i + \gamma I)\|)\|x\| + (1+\eta_0)\rho_t(x, t) + \epsilon)\text{Tanh}(\sigma_i) \quad (23)$$

*Remark 7.* Taking the norm of (20) and applying the results from Remark (5) yields  $\|U_a\| \leq \|S_i \bar{A}_i\| \|x\| + \|\gamma S_i\| \|x\| + \|\rho_t\| \|\text{Tanh}(\sigma_i)\|$ . Since  $\|\text{Tanh}(\sigma_{ij})\| \leq 1$  for all  $\sigma_i$ ,  $\|x\| \leq B_x$ , we obtain the following explicit upper bound on the norm of the demanded torque input  $\|U_a\| \leq \|S_i \bar{A}_i\| B_x + \|\gamma S_i\| B_x + \|\rho_t\|$ .

## 4 Simulation Studies

Let us consider a wheeled mobile under-actuated manipulator shown in Fig. 2. The following variables have been chosen to describe the vehicle (see also Fig. 2):



**Fig. 2.** The wheeled mobile manipulator in the simulation

$\tau_l, \tau_r$ : the torques of two wheels;  $\tau_1$ : the torques of joint 1;  $\theta_l, \theta_r$ : the rotation angle of the left wheel and the right wheel of the mobile platform;  $v$ : the forward velocity of the mobile platform;  $\theta$ : the direction angle of the mobile platform;  $\omega$ : the rotation velocity of the mobile platform, and  $\omega = \dot{\theta}$ ;  $\theta_1$ : the joint angle of the under-actuated link;  $m_1, I_1, l_1$ : the mass, the inertia moment, and the length for the link;  $r$ : the radius of the wheels;  $l$ : the distance of the wheels;  $l_G$ : the distance between the wheel and joint 1;  $m$ : the mass of the mobile platform;  $I$ : the inertia moment of the mobile platform;  $I_w$ : the inertia moment of each wheel;  $g$ : gravity acceleration.

The mobile under-actuated manipulator is subject to the following constraint:  $\dot{x} \cos \theta - \dot{y} \sin \theta + \dot{\theta} l_G = 0$ . Using Lagrangian approach, we can obtain the dynamic model with  $q = [\theta_l, \theta_r, \theta_1]^T$ . For the page limit, the detail information of dynamics is omitted. As discussed in Section II, we set the fully operational configuration represented by OOO while three possible fault configurations can occur: AAP, APA, and APP, where *A* represents actuated joints and *P* represents failed joints. For example, if we find that a fault occurs in  $\tau_{\theta_1}$ , then the fault configuration to validate the proposed methodology is the AAP configuration, the configuration AAP means that the two wheels are active and the joint

is passive. We consider a workspace with a positioning domain which range from  $q(0) = [0^\circ, 0^\circ, 0^\circ]$  to  $q(T) = [20^\circ, 20^\circ, 20^\circ]$ , with the velocities set to  $1^\circ/s$ , and use 2 sectors of position in each joint, denoted as I( $0^\circ \sim 10^\circ$ ) and II( $10^\circ \sim 20^\circ$ ), to map the mobile manipulator workspace. The linearization points with respect to I and II are chosen as  $5^\circ$  and  $10^\circ$ , respectively. Then, according to above theories, 8 linearization points with 32 modes are found. The feasible solutions are difficult to find because of the 32 modes, so we select the 8 typical modes to do simulation. which are shown in Table 4. There exist 8 modes for the above

Mode	Joint Status	Mode	Joint Status	Linearization Section		
				$\theta_r$	$\theta_l$	$\theta_1$
1	APA	5	AAP	I	I	I
2	APA	6	AAP	II	I	I
3	APA	7	AAP	II	I	II
4	APA	8	AAP	II	II	II

fault-tolerant example, which means a  $8 \times 8$  dimension transition rate matrix  $\Pi$  is needed. So the  $\Pi$  is defined as

$$\Pi = \begin{bmatrix} -0.72 & 0.15 & 0.22 & 0.21 & 0.14 & 0 & 0 & 0 \\ 0.2 & -0.7 & 0.2 & 0.2 & 0 & 0.1 & 0 & 0 \\ 0.16 & 0.22 & -0.68 & 0.2 & 0 & 0 & 0.1 & 0 \\ 0.22 & 0.3 & 0.2 & -0.82 & 0 & 0 & 0 & 0.1 \\ 0 & 0 & 0 & 0 & -0.78 & 0.26 & 0.26 & 0.26 \\ 0 & 0 & 0 & 0 & 0.26 & -0.78 & 0.26 & 0.26 \\ 0 & 0 & 0 & 0 & 0.26 & 0.26 & -0.78 & 0.26 \\ 0 & 0 & 0 & 0 & 0.26 & 0.26 & 0.26 & -0.78 \end{bmatrix} \quad (24)$$

The the system parameters are chosen as  $m = 10kg$ ,  $m_1 = 2kg$ ,  $l_1 = l_2 = 1.0m$ ,  $I_m = 2kgm^2$ ,  $I_w = 2kgm^2$ ,  $I = 1kgm^2$ ,  $I_1 = 1kgm^2$ ,  $r = 0.5m$ . Then from (11), we get the linearized system matrices  $A_i, B_i$  for  $i = 1, 2, \dots, 8$ . For the page limit, we omit these in this paper. Assume that  $\Delta A_i, \Delta B_i$  for  $i = 1, 2, \dots, 8$ , are given by

$$\Delta A_i = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \lambda_1 \cos(u_1) & 0 & 0 \\ \lambda_2 \cos(x_1) & 0 & 0 & 0 \end{bmatrix}, \Delta B_i = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \lambda_3 \cos(x_1) \end{bmatrix}.$$

where  $\lambda_i$  are unknown but bounded as  $|\lambda_1| \leq 0.00002$ ,  $|\lambda_2| \leq 0.00002$  and  $|\lambda_3| \leq 0.00001$ . So we can obtain  $\rho A_i = 0.00002$ ,  $\rho B_i = 0.00001$ , for  $i = 1, 2, \dots, 8$ , and  $\rho_f(x, t) = 1$ . Via LMI optimization with the data  $A_i, B_i, \Delta A_i, \Delta B_i, \rho A_i$  and  $\rho B_i$ , we can get the following solution  $X_i (i = 1, 2, \dots, 8)$ . So we can obtain the solution of  $S_i$ , for  $i = 1, 2, \dots, 8$ . Torque disturbances  $D(t)$  are introduced to verify the robustness of the controllers



$$\begin{bmatrix} d_r(t) \\ d_l(t) \\ d_1(t) \end{bmatrix} = \begin{bmatrix} 0.023 \sin(4t) \\ 0.007 \sin(3t) + 0.009 \cos^2 t \\ 0.015 \cos(5t) \end{bmatrix}$$

The disturbance is turned off after the fault introduction in corresponding joint or wheel. By using (20) and (23), we can obtain the SMC laws. We assume that  $\lambda_1 = 0.00002 \sin(10\pi t)$ ,  $\lambda_2 = 0.00002 \sin 20\pi t$ ,  $\lambda_3 = 0.00001 \sin 30\pi t$ ,  $\gamma = 1$  and  $\epsilon = 0.01$ . Consider the uncertain system (12), the initial condition we used for simulation is  $x_0 = [0.1, 0.1, 0, 0]^T$ .

The system switching pattern among the 8 modes randomly during operation. From Fig 3 we can see: first the system switches mode 1 to mode 3, then the system switches mode 3 to mode 1, and the system switch mode 2 to mode 1 finally.

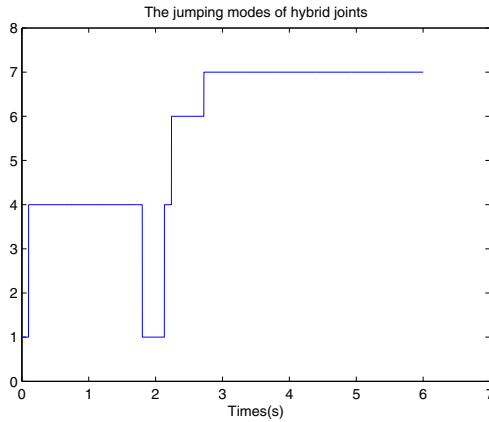


Fig. 3. Jumping

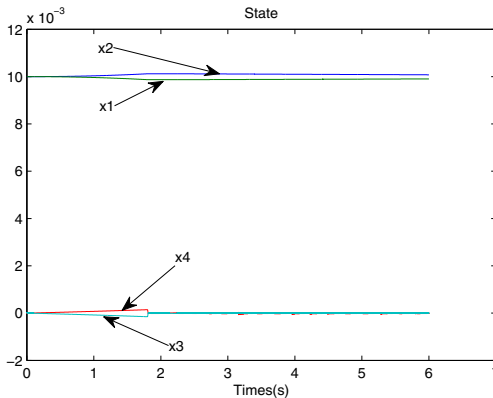


Fig. 4. States

## 5 Conclusion

A Markovian switching model was developed for a mobile manipulator with two independent wheels and  $n_a$  joints. The uncertainty of the transition rate matrix is considered in an element-wise way. We have presented sufficient conditions on the existence of mode-dependent feedback controller, which guarantees the robust stochastic stability of the closed loop system for all admissible uncertainties including the unprecise transition rate.

## References

1. Li, Z., Ming, A., Xi, N., Shimojo, M.: Motion control of nonholonomic mobile underactuated manipulator. In: IEEE International Conference on Robotics and Automation, pp. 3512–3519 (2006)
2. Li, Z., Ming, A., Xi, N., Gu, J., Shimojo, M.: Development of hybrid joints for the complaint arm of Human-Symbiotic Mobile Manipulator. *International Journal of Robotics and Automation* 20(4), 260–270 (2005)
3. Bergerman, M., Lee, C., Xu, Y.: A dynamic coupling index for underactuated manipulators. *Journal of Robotic Systems* 12(10), 693–707 (1995)
4. Spong, M.W.: The swing up control problem for the Acrobot. *IEEE Contr. Syst.* 15, 49–55 (1995)
5. Li, Z., Ge, S.S., Ming, A.: Adaptive robust motion/force control of holonomic-constrained nonholonomic mobile manipulator. *IEEE Trans. System, Man and Cybernetics, Part B: Cybernetics* 37(3), 607–617 (2007)
6. Mariton, M.: *Jump Linear Systems in Automatic Control*. Marcel Dekker, New York (1990)
7. Boukas, E.K.: *Stochastic Hybrid Systems: Analysis and Design*. Birkhauser, Boston (2005)
8. Choi, H.H.: LMI-Based Sliding Surface Design for Integral Sliding Mode Control of Mismatched Uncertain Systems. *IEEE Trans. Automatic Control* 52(4), 736–742 (2007)
9. Li, Z., Kang, Y.: Dynamic Coupling Switching Control Incorporating Support Vector Machines for Wheeled Mobile Manipulators with Hybrid Joints. *Automatica* 46(5), 785–958 (2010)
10. Li, Z., Yang, Y., Wang, S.: Adaptive Dynamic Coupling Control of Hybrid Joints of Human-Symbiotic Wheeled Mobile Manipulators with Unmodelled Dynamics. *International Journal of Social Robotics* 2(2), 109–120 (2010)

# Timestamp Based Predictive Robot Control System over Real-Time Industrial Ethernet

Huasong Min<sup>1,2</sup>, Tianmiao Wang<sup>2</sup>, Hongxing Wei<sup>2</sup>, and Zhenglin Liu<sup>1</sup>

<sup>1</sup> School of Computer, Wuhan University of Science and Technology, Wuhan, 430063, China

<sup>2</sup> School of Mechanical Engineering and Automation, Beihang University,  
Beijing, 100191, China  
mhuasong@buaa.edu.cn

**Abstract.** Real-time Industrial Ethernet will become the development trend of the robot control system. The paper describes main methods how to use industrial Ethernet for robot control system, put forwards some ideas to get a good real-time and low-jitter validation of the robot control system from the following six aspects: bus topology architecture, clock synchronization technology, switching technology, delay prediction algorithm, UDP protocol improving, prediction and compensation control. It gives a timestamp based predictive robot network control system model which combines the technology of clock synchronization, state estimation, delay prediction, predictive control, and the robotics network control system is simulated with Matlab/Simulink and Truetime.

**Keywords:** Real-time Industrial Ethernet, Robot Control System, Truetime.

## 1 Introduction

In recent years, Serial Bus has been widely applied in the complexity robot and modular robots. Robot control system based on serial bus has been rapid developed abroad. However, although the serial bus control system can control more of the shaft, but the requirement of network bandwidth and real-time bottlenecks encountered. In the serial bus control system, the reliability of the system will be reduced and the cost increased while the bus speed is improved.

With the development of computer software technology, network technology, embedded software/hardware technology, and industrial Ethernet technology, it has a strong feasibility and high research value to apply industrial Ethernet to robot control system and design complex robots, modular robots and multi-robot control system through embedded hardware and software.

Industrial Ethernet technology was originated from the traditional Ethernet technology. Because of Ethernet with a high communication bandwidth, it can full fill the requirement of communication speed in complex robot control system. But it use the medium access mechanism with collision detection Carrier Sense Multiple Access (CSMA/CD) and the node with Binary Exponential Backoff(BEB) algorithm for processing conflict, which has the defect of uncertainty of the communication delay as a major obstacle in real-time industry control application. In order to meet the

response time less than 5ms in application, various industrial Ethernet protocols organizations have given some solutions to enhance real-time industrial Ethernet. These include China's EPA and other five types real-time Industrial Ethernet [1] (EtherCAT, Ethernet Powerlink, PROFINET, Modbus-IDA, Ethernet/IP), become the six major competitors.

## **2 The Research of Applying Ethernet Technology to Robot Control System**

Applying Ethernet in control area originated from Goldberg who first set up Internet-based remote control robot system--Mercury Project in the University of California, 1994, which bradford space machine telescope is similar to. Users can operate the telescope online through the Internet [2], Internet-based remote control was born. Their pioneering work provided us with a new idea to use the Ethernet. Since then, the research and application of Ethernet prospered. The scholars of USA Michigan State University and Japan Nagoya University had put forward an enhanced Internet-based robot in 2002, which presented modeling and analysis tools to the network delay and uncertainty [3]. J Sanchez F. Morilla discussed the Internet-based control laboratory [4], the laboratory uses client-server mode, the server is located in the vicinity of the controlled object and passes over control signals of the client to the controlled object, according to output volume of the controlled object received, the client located in the terminal calculates the control and transfer it to the server to form a closed loop control system. Current research focus is how to improve the real-time of Ethernet. In the research of G.P. Liu, S.C.,and Chai, et al. [5], for the fixed delay case, a stability criterion is derived for the closed-loop system and is expressed by the stability of a single polynomial, for the random delay case, the system stability criterion is much more complicated and is expressed by the stability of four Kharitonov polynomials. These opened up new ideas for studying robot control network.

The research of current Ethernet-based robot control system usually is improving the algorithm under the TCP/IP framework, does not involve the network layer and network interface layer of the Ethernet. With the integration of control network and communication network, the industry has released a variety of real-time industrial Ethernet standards such as EtherCat, Ethernet Powerlink, PROFINET, Modbus-IDA, Ethernet /IP, EPA, etc., industrial Ethernet has made a good use. But for the robot control system, because of its complexity and higher real-time requirements, the research of Industrial Ethernet in robot control systems is still in the initial stage.

## **3 The Main Methods of Industrial Ethernet Applied to Robot Network Control System**

Robot control system needs an entirely open architecture and needs interconnection with the existing network. Applied the real-time industrial Ethernet to robot network control system, the current research can be divided into two directions. One is the complete transformation of industrial Ethernet, develop a dedicated network protocol and interface chip directly for robot, which can be called as "Robot Control over

Ethernet". Another idea is to transform the upper communication protocol based on the present real-time industrial Ethernet, using control means or adding new technology to make delay a predictable and controllable in the situation of network delay existing and unavoidable, to meet the requirements of the robot distributed network control, which can be named as "Robot Control with Ethernet".

This paper discusses on the point of how to apply the industrial Ethernet to robot control, proposes ideas from six areas to improve the performance, there are the topological structure of robot network control system, clock synchronization technology, network switch technology, delay analysis, delay prediction, improvement of UDP protocol, prediction and compensation control algorithm.

Robot control bus topology can not be oversimplified, must meets the diverse needs of the robot form. Currently, there are many topology structures which are common used for network control, such as star topology, bus type, tree topology and other. But for modular, networked robot control system, must according to the needs for different applications and requirements, adopt appropriate network topology. For examples, for a simple reptile robot, its control system architecture can choose to star topology, but for complex humanoid robot and multi-axis industrial robot, it maybe needs to adopt bus-type topology, and for the network of multiple robots, even Internet-based connection, will be the interconnection between multiple heterogeneous networks.

Since Ethernet adopt the "best-effort" transmission mechanism, so there is always a lack of timing synchronization, in industrial or other high real-time demanding applications. The synchronization accuracy of Network Time Protocol (NTP) 1992 version is within 200 $\mu$ s, but still can not meet the requirement of measurement instruments and real-time industrial control. With the support of the United States Institute of Standards and Technology (NIST), which is under the IEEE Instrumentation and Measurement Committee, the specification drafted by the committee, was adopted by the IEEE Standard Commission at the end of 2002, as the IEEE 1588 standard [6].

In 2008, IEEE 1588 v2 version released, and UTC (Coordinated Universal Time) synchronization, time synchronization accuracy can be achieved even nanosecond accuracy level. Adopt the IEEE1588 to robot network control, particularly high-speed real-time motion control system, the PTP (Precision Time Protocol) protocol will play a key role, in real-time scheduling and network delay measurement, combined with advanced scheduling algorithms to improve the stability and real-time.

### 3.1 Switching Technology and Delay Analysis

If the robot control system is more complex and have arduous communication task, we can use conflict domain partition technique, to divide a shared local network into some separated collision domains, each of the collision domain connected with full-duplex switches and to reduce collisions and errors caused by the CSMA/CD transmission mechanism. This method can avoid collisions as much as possible, increase the certainty of the system.

There are three methods to analysis network delay. This first is fixed delay method, by adding caching mechanism to the network nodes. This method artificially increases the delay. The second method depends on that network delay follows a distribution law. The last one considers that latency is changing, but the delay is bounded.

This paper agrees with the third method, at the  $K$  times sampling time between the two modules communicating, the delay is described in the formula 1.

$$\tau_k = T_{def} + T_{csma} + T_{trans} + T_{encode} + T_{decode} + T_{wait} \quad (1)$$

$T_{def}$ : Channel state detection time.  $T_{csma}$ : the back off time when detected conflict.

$T_{trans}$ : Data transfer time, which depends on the length of the packet.  $T_{encode}$ : Time of sending data to be packed.  $T_{decode}$ : Decoding time of received packet.  $T_{wait}$ : waiting time for the data received in the buffer to be obtained or in the sending buffer to be sent.

The uncertain delay depends on  $T_{csma}$ , and if using full-duplex switched Ethernet can greatly reduce the time for conflict resolution.

This paper discusses the network control on the premise that delay is bounded, delay can not be infinite, and it has maximum delay  $T_{max}$  and minimum delay  $T_{min}$ . For 100BASE-T switched Ethernet, for example, one bit data transmission time can be calculated as formula 2.

$$T_t = \frac{1}{100 * 10^6} = 0.01 \text{ us} \quad (2)$$

Signal transmission rate in the cable is  $T_v = 1.9 \times 10^8 \text{ m/s}$ . In this paper, assuming the maximum cable length as 100m, according to the network control system of a robot control 20 motors (For example, in our dog robot project, we use 16 motors to consist of a network), each motor contains control information and feedback information, and data length by a maximum of 18 bytes, at one times transmission the packet length is 384 bytes (data length is  $20 * 18 = 360$  bytes, the other overhead 18 bytes, preamble occupies 8 bytes). Minimum delay for the buffer when the switch no data storage, data directly, data frame delay is calculated as formula 3.

$$T_{min} = T_t + T_{trans} = 384 * 8 * T_t + 100/T_v = 30.72 + 0.52 = 31.24\text{us} \quad (3)$$

$T_t$  is the frame send delay,  $T_{trans}$  is the signal transmission delay in network.

Maximum delay occurs at the time when the switch buffer is full fill with the data must send to a node. The maximum delay equals to minimum delay plus waiting delay. According to the algorithm of the reference paper [14], the sampling period is 1ms, the maximum delay of 100M switched Ethernet is described in formula 4.

$$T_{max} = 31.24 + 9 * (96 + 576) * 0.01 * N = 31.24 + 60.48 = 91.72 \text{ us} \quad (4)$$

Where  $N$  is the cache nodes numbers of network switch, according to the maximum number 9 nodes send data to the one same node, sending data from 9 nodes will be cached.

From this we can see that if we adopt 100M switched network, in the 1ms sampling period of control, the delay is not widespread, and can be applied to multi-axis robot real-time control, and multi-robot group real-time control. If we increase the network bandwidth such as increase to 1000BASE-T, also can significantly reduce the delay. From the smallest to the largest delay analysis of switched Ethernet networks proved that switch network can be used for robot strict real-time control.

### 3.2 Delay Prediction

Delay measurement provides the basis for the next time delay estimation. Delay prediction algorithm is through the last period delay values to predict the value of the next period delay. SUN Li-ning [7] et al.(2004) use linear neural network and RBF neural network to predict delay in the Ethernet. Liu D.Du, Y.H.Zhao [8] et al.(2004) use Elman neural network for an intelligent network latency prediction, and Smith Delay Compensator for delay compensation. In Wang Qingpeng et al. paper(2001) [9], Round-trip delay for the Internet has been tested, and then in the TCP protocol in the round-trip delay estimation algorithm is proposed based on a new round-trip delay estimation algorithm. CAN bus time delay estimation methods were discussed in Wei zhen et al.(2003) [10]. Nikolai Vatanski et al. (2009) [11] summarize the effect of delay and the methods of removing delay effect in network control system. Summary of various time delay estimation methods, there are three delay estimate methods which are used most widely. The first is fixed delay method, in all the transmission process, the network delay is considered as a constant. The second is linear weighting factor prediction method, this method is through over a period of time delay values, through a linear model to estimate the value of the next moment of delay. The last is neural network prediction method, the best advantages of neural network prediction method is that it has the capacity of an arbitrary non-linear approximation. BP neural network algorithm and the RBF neural network algorithm is a common neural network algorithm to estimate delay.

Although network delay of 100M switched Ethernet network is small, but for high-speed motion control, even dozens of microseconds of the synchronization deviation will result in large jitter. The improved timestamp based delay prediction algorithm will be described in section 3.4.

### 3.3 Improvement of UDP Protocol

The common protocols in network transport layer are TCP and UDP. The timeout-retransmission mechanism of TCP protocol take up a lot of resources, seriously affect the system's real-time, with the contrary, UDP protocol do not need to set up or remove the connection, which meets the requirements of real-time control system, but the UDP is a connectionless transport mechanism, and can not guarantee orderly and reliable when data reached the destination, it will happen packet loss, duplicate packets received, out of order, etc., so it need algorithm to ensure the reliability of data transmission. In order to ensure system reliability, real-time requirements, the following work is done to improve UDP protocol performance.

(1) First, good data transfer format designed to ensure that the same control information and feedback information is encapsulated in the same packet, that's the single packet transmission, so as to avoid the sequence disorder of the same information packet in receiving data which result from the uncertainty of delay.

(2) Set up UDP buffer and the buffer size should take into account both the requirements of real-time, not too big, and the problem of overflow.

(3) In order to prevent packet loss, the frequency of master and slave controller should be different. The frequency of send data in slave controller is designed  $N$  ( $N = 2, 3, \dots$ ) times the master controller in this paper, that is to say, the master

controller can receive data more than one copy (at least two) from the slave controller, which can prevent the problem of the loss-packet cannot be re-transmitted, due to the transmission cycle time is too short.

### 3.4 Prediction Robot Control over Industrial Ethernet

Network control brings changes with the structure of motion control, broadens the scope of application and makes the distributed control to be possible. The simple model of a general network control is shown in Figure 1.

It consists of two parts, network closed-loop system and local closed-loop system. Controller sends the position reference (or current, speed) to actuator at the every start time of bus communication. Actuator use the position reference to control the motor (controlled objects) closed-loop position or the current or speed. At the same time the actuator feedback data to the controller periodically. Controller based on feedback data to calculate the new position to send it and into cyclical repetition.

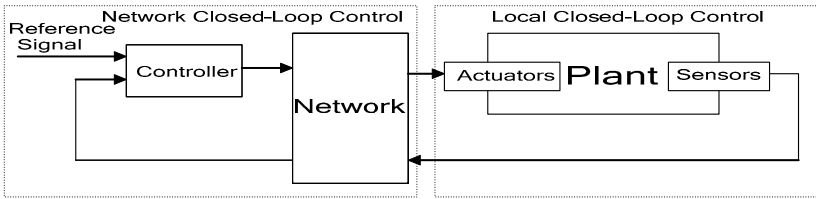


Fig. 1. General Network Double-loop Control System

The compensation control of Smith predictor is concerned widely, it takes the network delay as a lag element of time uncertain, adds Smith predictor into the loop, and uses it to predict the system state with delay, and then make compensation control. Predictive control build the controlled object model at the master controller end, and then based on the prediction delay, simulating the state of the controlled object in local model.

The robot network control model of this paper works on the premise that master-slave controller clocks are synchronized by IEEE 1588 PTP protocol, there is high-precision clock synchronization between master and slave. As shown in Fig.2, 't' is now time,  $u(t)$  that is now time to send the feedback information from the controller via UDP protocol. Master receives the data, using the local clock and compare with the data time stamp, obtained feedback delay  $\pi$ . The  $\pi$  and  $u(t)$  are used in the model of the slave controller, after operations, received the state of slave controller,  $u(t + \pi)$ , and at this time through the feed-forward delay prediction algorithm to be estimated delay is  $\epsilon$ , In the master controller after operation with  $\epsilon$  and  $u(t + \pi)$ , get the control information of the next communication cycle,  $y(m)$ , after put time stamp on  $y(m)$ , then send by the UDP, the slave controllers receive the control message, use this information to refresh the control parameters of the controlled object, and return sensor data and time stamp packages of the controlled object, then send via UDP. A new communication cycle again.



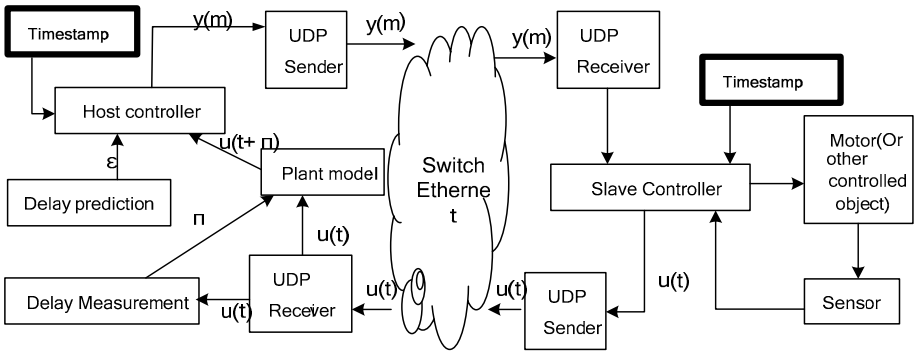


Fig. 2. Timestamp Based Predictive Robot Control System model over Real-time Industrial Ethernet

### 4 Simulation of Robot Control over Ethernet without Time Synchronization and Delay Predictive

Because of Ethernet high-speed throughput, Ethernet can be used in industrial robot control. The purpose is controlling the communication between multiple motors and sensors, and achieving the motion control of more than 6-axis robot or multi-robot. Robot control object model should at least establish multiple motor model of network interconnection. But for network interconnection system, we can only simulate the most extreme delay motor. Therefore, the simulation model of this paper is listed only one motor node whose delay is artificially set to largest.

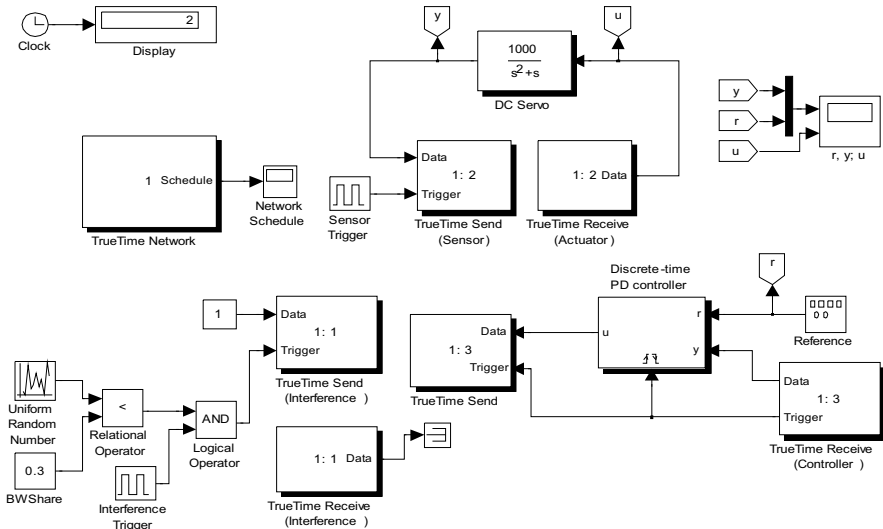
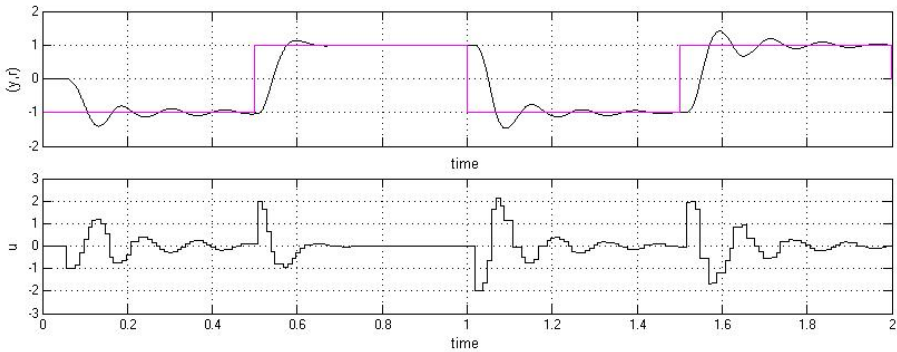


Fig. 3. The model of robot network control without delays predictive

This simulation used one Truetime Network module, three Truetime send module and three receive module. The network type in Truetime Network module is set to Switched Ethernet, bandwidth of 100MB/S, the minimum frame is 3072b (384Byte), packet loss rate of 0.03. A Discrete-time PD controller module which is responsible for receiving from the signal source and sensor feedback signals and processing these signals and its clock offset is 10ms, A Truetime Send module as a sensor node, Truetime receive module as an actuator node, and the signal source is square wave. The model of robot network control without delays predictive shown in Figure 3.

The simulation results are shown in Figure 4.



**Fig. 4.** Output ( $y$ ) and controller ( $u$ ) results of simulation of Shared Ethernet motor control

From Figure 4, in a switched Ethernet, there also exists jitter of motor control. The reason is uncertainty of network delay and packet loss occurrence. The control command reach the actuator after a period of delay time, and status data returned from sensor is not the “now time” state, therefore the adjustment data calculated by PID does not obtain the desired results.

## 5 PTP Timestamp-Based Predictive Control of Robot Network

If want to get good result, network control must be improved, where we propose a timestamp based on the PTP model predictive control to improve robot control based on industrial Ethernet. By calculating the data transmission in Ethernet frame delay to estimate the next time delay, the controller side of the object model, model simulation of the accused in the local state of the object, and its compensation control combined with delay prediction and control.

The improved control model works on the premise that master-slave controller clocks are synchronized by IEEE 1588 PTP protocol, there is high-precision clock synchronization between master and slave.

Compared with the traditional NTP synchronization, time-stamp join in the application layer, PTP protocol move the stamping time down to the above of physical layer, following the MAC layer, this requires hardware support, of course, it can also

be stamped at the application layer by software. In this paper, the controller node of the simulation model, the clock offset is set to 100us (Reference to the accuracy range of PTP protocol software synchronization), that means the clock synchronized through Precise Time Protocol.

After adopt technology of Ethernet switching, Delay Prediction, State Prediction, etc., the improved simulation model shown in Figure 5.

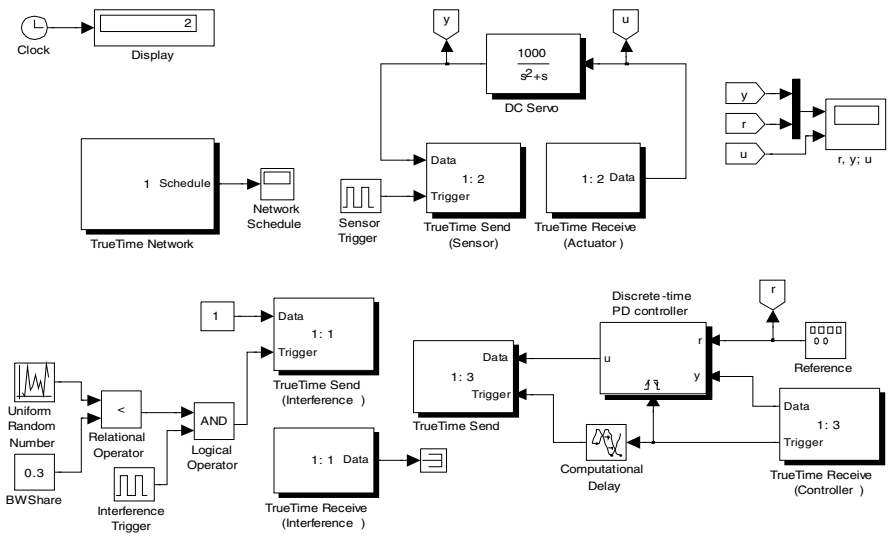


Fig. 5. Improved model of switched Ethernet motor control

Linear delay prediction cannot predict the delay value accurately, only for little change of network delay, the greatest advantage of neural network algorithm is that it can approximate nonlinear value at will, but they often require significant historical data to train the neural network, this method need large amount of calculation and often at the initial stage can not reach accurate results because of lack of time. Integrated linear delay prediction with neural network delay prediction, the paper proposed a new algorithm by using a linear prediction based on time stamp, and combined BP neural network algorithm, the specific ideas described as follows.

Before propose delay prediction algorithm, we do some assumptions.

(1) Only use a single packet transmission between controller, and the maximum network delay is less than a bus cycle of delay by the previous analysis. As already discussed in this paper, it is easier to meet this assumption.

(2) Each node in the system has been synchronized by IEEE1588 protocol.

So, the improved algorithm of this paper is divided into two stages, and the flow-chart shown in Figure 6.

(1) Linear prediction stage. Delay prediction at this stage use linear prediction, because in the initial stages of the delay did not get a lot of data for BP neural network prediction method. At this stage, the prediction algorithm as follows.

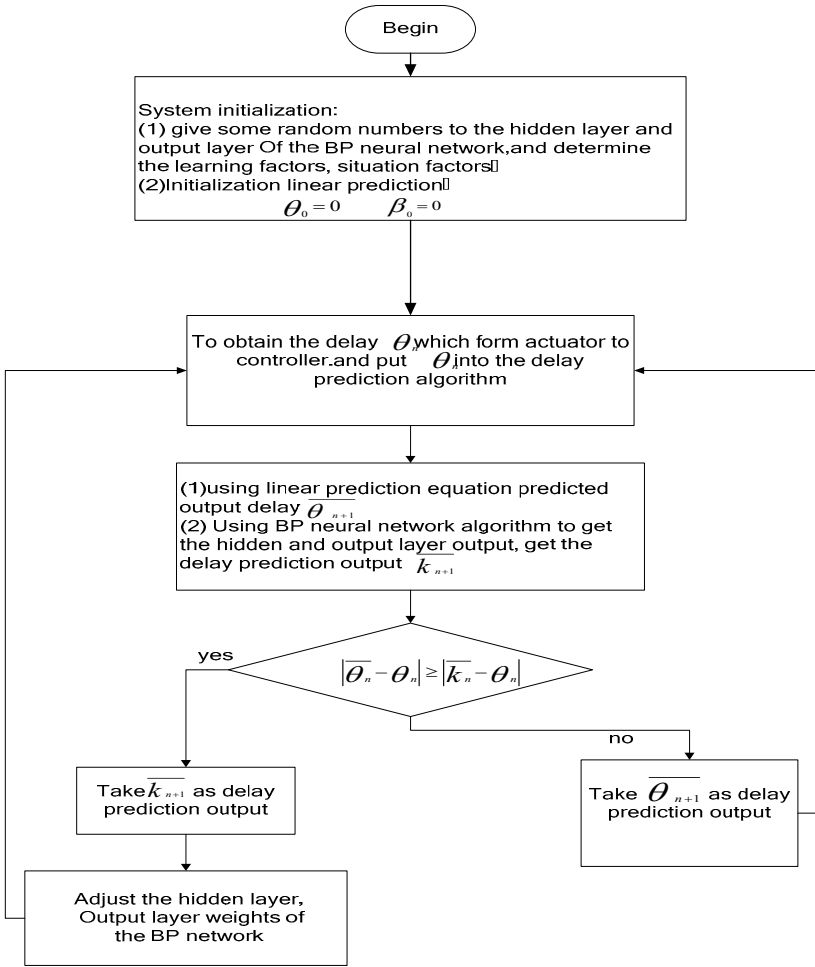


Fig. 6. Flow chart of delay prediction algorithm

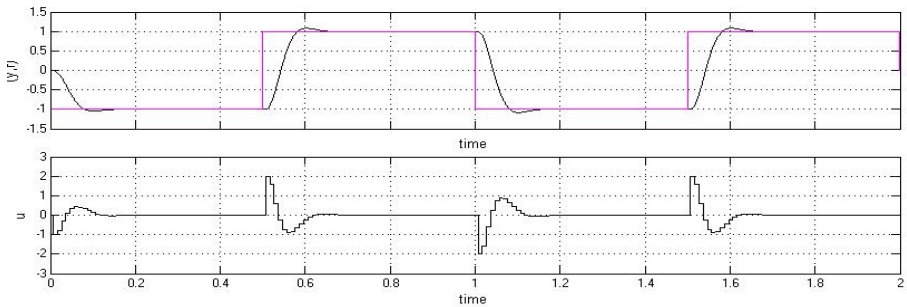


Fig. 7. Output (y) and controller (u) results of improved network control model

Clock synchronization, the controller sends control information to the actuator, the data is sent to affix the stamp, when the actuator received message, using the local clock and time stamp by comparing the data from controller to actuator delay.

At this stage BP neural network algorithm is also enabled, but the result is not used to predict the output, mainly through a period of time training to let all levels connection value of BP neural network to be trained.

(2) BP neural network prediction stage. By the training of the previous stage, when the BP neural network prediction of the delay offset is less than the linear prediction offset, set the BP neural network prediction value to output, and linear prediction to stop working.

The idea of the algorithm is based on compare the linear prediction error with the BP neural network prediction offset, which is smaller, then to chose the smaller offset method to prediction delay, and output.

In this simulation the motor transfer function is  $G(s) = \frac{1000}{s^2 + s}$ , sampling period of 0.002s, clock offset set to 100us (assuming the system synchronized via IEEE1588 algorithm), the Truetime Network module protocol selected for Switch Ethernet, the other parameters are same as the before model shown in Figure 2.

After delay compensation and predictive control, the system output shown in Figure 7.

Compare Figure 7 with Figure 4, the better real-time, low-jitter of the system is achieved with improved model.

## 5 Conclusion

This paper gives an timestamp based predictive robot network control model which combines the technology of clock synchronization, state estimation, latency prediction, predictive control, and the robotics network control system is simulated with Matlab/Simulink and Truetime. Compared with the model without clock synchronization and delay predictive, the improved model achieved better real-time, low-jitter of the system.

**Acknowledgments.** This work is supported by the 863 Program of China (2009AA043901 and 2009ZX04013-012-01).

## References

1. Falser, M.: Real-time Ethernet industry prospective. *Computing & Control Engineering Journal* 93(6), 1118–1129 (2005)
2. Goldberg, K., Mascha, M., Gentner, S.: *Desktop teleportation via the World Wide Web*. IEEE, Piscataway (1995)
3. John, E.F., Cox, M.J.: Remote control and robots, an internet solution. *Computing & Control Engineering Journal* (February 1996)
4. Sanchez, J., Dormido, F., Aranda, J., Ruiperez, P.: *Virtual and Remote Control Labs Using Java: A Qualitative Approach*. IEEE Control System Magazine (April 2002)

5. Liu, G.P., Chai, S.C., et al.: Networked predictive control of systems with random delay in signal transmission channels. *International Journal of Systems Science* 39(11), 1055–1064 (2008)
6. ANSI/IEEE Standard 1588-2002, IEEE standard for a precision clock synchronization protocol for networked measurement and control systems (2002)
7. Sun, L.-n., Xie, X.-h., Zhang, F.-f., et al.: Research on neural network based prediction algorithm of time delay. *Robot* 26(3), 237–240 (2004)
8. Liu, D., Du, Y.H., Zhao, Y., et al.: Study on the time-delay of Internet-based industry process control system. In: *Proceedings of the 5th World Congress on Intelligent Control and Automation*, Hangzhou, pp. 1376–1380 (2004)
9. Wang, Q.-p., Tan, D.-l., Chen, N.: Network time delay test and analysis in the control of internet based telerobotics. *Robot* 23(4), 316–321 (2001)
10. Zheng, W., Changhong, L., Jianying, X.: Online delay-evaluation control for networked control systems. *Control and Decision* 18(5), 545–549 (2003)
11. Vatanski, N., Georges, J.-P., et al.: Networked control with delay measurement and estimation. *Control Engineering Practice* 17, 231–244 (2009)

# Design and Experimental Research on a Fast Zero-Finding Algorithm Based on the Incremental Angle Encoder

Zhonghua Miao<sup>1</sup>, Xuyong Wang<sup>2</sup>, Chengliang Liu<sup>2</sup>, and Xiaojin Zhu<sup>1</sup>

<sup>1</sup> School of Mechatronics Engineering and Automation, Shanghai University

<sup>2</sup> School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China  
Zhonghua\_miao@sjtu.edu.cn

**Abstract.** The incremental angle encoder takes the current initial position as the zero point when switching-on, therefore, it's necessary to establish the mechanical origin of the system in some angle positioning systems, i.e., zero-finding. Addressing this issue, a novel zero-finding algorithm and its realization are proposed based on the incremental encoder with distance-coded marks. The experimental results show that the proposed algorithm is feasible, which can not only quickly establish the mechanical origin of the system without require a full circle rotation, but also greatly improve the accuracy, the robustness and the ability to anti-interference.

**Keywords:** Zero-finding algorithm, Incremental angle encoder, Mechanical origin, Sensor, RON786C.

## 1 Introduction

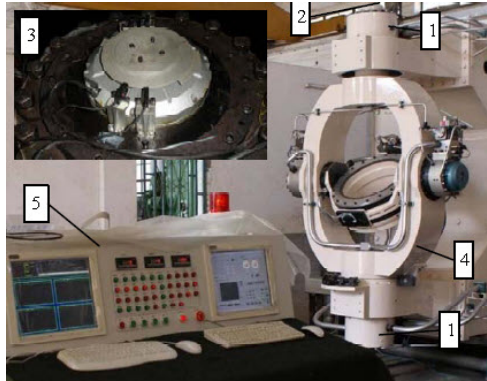
The incremental angular encoder is widely used in the fields of precision angle measurement and positioning control for its more accurate, price and market decision [1-5]. However, it takes the current position as the zero point when switching-on, but actually it isn't the zero point expected. In most cases, therefore, it's necessary to find out the mechanical origin of the system to obtain the absolute position especially for the incremental displacement measurement system. To address the problem on how to achieve absolute positioning by use of the incremental encoder, many research efforts on zero-reference mark have been reported. In [6], the encoder started to count pulses when traversing the zero-reference mark after switching-on each time, then it would supply a unique position value at certain position even if switching-on at different positions. Literature [7] also introduced a practical method to achieve absolute positioning based on the incremental encoder by virtue of Siemens PLC (Programmable Logic Controller).

The above methods are efficient and practical but the drawback is that the encoder may require a rotation up to nearly 360° in some cases, and fail to find a reference mark in some equipment which has a limited range of the working angle [8]; on the other hand, the zero-finding precision of the encoder with one reference mark is greatly decided by the sensitivity of special position of the system, and it is hardly

immune to the errors caused by the disturbance. Therefore, in order to speed up and simplify such “reference runs” and improve the robustness, a novel zero-finding algorithm is proposed in this paper. After the operation of the zero-finding, the mechanical absolute origin can be rapidly ascertained after traversing one nominal incremental angle, and each absolute position on the circumference is established in the incremental angle encoder.

## 2 Setup and Problem Statement

Fig.1 shows the photograph of the three-axis hydraulic servo turntable which is mainly composed of three frames (outer frame, middle frame and inner frame), five hydraulic motors (up and low motors, right and left motors, hollow motor) and digital servo valves (MOOG Company). The outer frame of the turntable machine is driven by the up and low hydraulic motors, the middle frame is driven by the right and left hydraulic motors, and the inner frame is driven by the hollow motor. The PCI-6601 card is used to acquire the angle signals generated by the incremental encoders named RON786C (HEIDENHAIN Company) which is mounted on the ending of the right and the up motor, respectively.



**Fig. 1.** Photograph of the experimental setup. (1) the hydraulic motors of the outer frame, (2) the incremental encoder mounted on the ending of the motor, (3) the amplification of the incremental encoder, (4) the outer frame of the three-axis turntable, (5) the table for data acquisition and system control.

Because the three-axis turntable machine is a precision trajectory tracking and position orientating system, which must meet the following conditions: ① The absolute positioning is required, i.e., the position value sent from the encoder must be put into a one-to-one correspondence with the physical rotary position on a circumference. No matter wherever the angle encoder is switching-on, the unique position value must be indicated on a certain position. ② The turntable must stop at the mechanical origin after switching-on each time, and await orders.

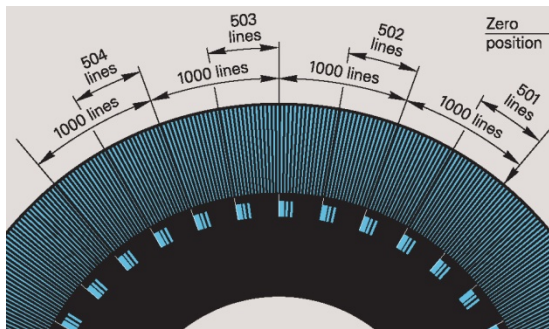


As we all know, it's easy for the absolute angle encoder, but difficult for the incremental angle encoder because it's based on the incremental measurement principle. Because of the uncertainty of the position where commenced, it's necessary to find out the mechanical origin of the system to obtain the absolute position. In this paper, it's expected to design a novel zero-finding algorithm to rapidly establish the mechanical origin by the incremental angle encoder. After the zero-finding, the incremental encoder will supply the absolute position value as that the absolute encoder could do. This will be a very interesting and valuable theme for engineers.

### 3 Preliminaries

In this section, the special coding mechanism of the incremental encoder with distance-coded reference marks is presented taking the RON786C as an example. Differing from the general incremental encoder with one reference mark, the RON786C features some distance-coded reference marks which are individually spaced according to a mathematical algorithm.

The RON786C encoder has altogether 36000 scaling marks around the circumference in the measurement of angle. Hereinto, the scaling mark on the aliquot position of  $10^\circ$  is the reference mark, and there is a sub-reference mark between two neighboring reference marks. The offset of two sub-reference marks is differentiated with the different positions of an angle. Both the position signals and the reference pulses will be sent out while traversing those reference marks. The marks are arranged as Fig.2.



**Fig. 2.** Schematic representation of circular scale of the RON786C sensor

The RON786C has three output channels named *A*, *B* and *R*. Every time traversing across a mark, a sinusoidal signal will be sent out from *A* and *B*, respectively. The amplitude *M* of the sinusoidal signal features  $0.6\text{V} \sim 1.2\text{V}$ , which typical value is  $1.0\text{V}_{pp}$ . When a reference mark (including sub-reference mark) is traversed across, a nearly triangular signal will be sent out from *R*. The amplitude *G* of the triangular signal is about  $0.2\text{V} \sim 0.85\text{V}$  when traversing across reference marks, however, which is about  $-0.04\text{V} \sim -1.7\text{V}$  when no reference marks are traversed across, listed as Fig.3. The signal from *R* is also called the reference signal in the following.

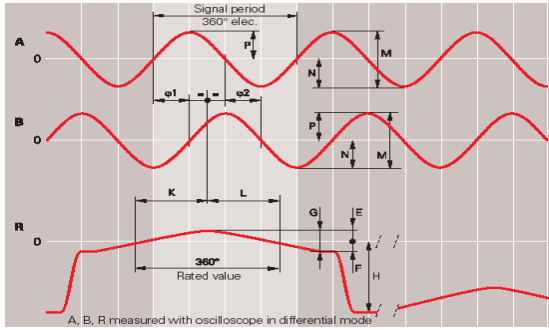


Fig. 3. Schematic representation of output signals of the RON786C sensor

### 4 Algorithm Design and Accomplishment

As stated in section 3, according to the particular coding mechanism, a zero-finding algorithm can be presented to deduce the zero reference position without traversing across the zero-reference mark. Three kinds of RC (*Reference Coordinates*) are defined as Fig.4: ① *RCSZE* - the RC of the Start Zero of Encoder, i.e., the reference coordinate under the switching-on position, ② *RCAZE* - the RC of the Absolute Zero of Encoder, ③ *RCMZS* - the RC of the Mechanical Zero of System. Generally, the *RCMZS* is what we need, and the position value used for close-loop control is the angle value related to the *RCMZS*.

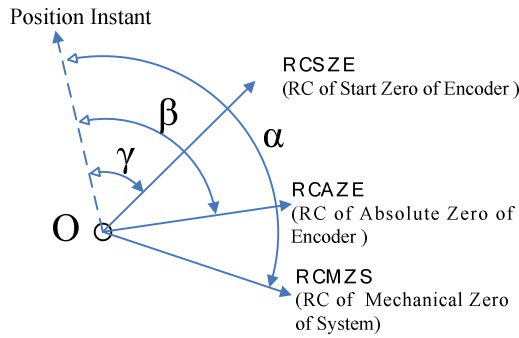


Fig. 4. Schematic representation of the reference coordinates

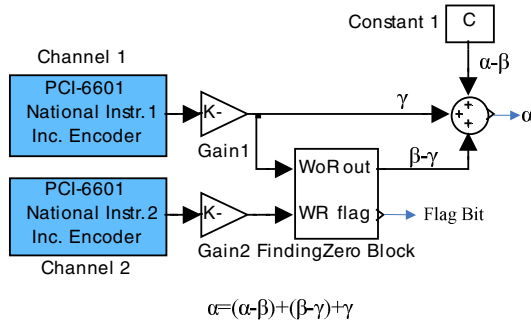
where,  $\alpha$  is the angle value in the *RCMZS* coordinate,  $\beta$  is the angle value in the *RCAZE* coordinate, and  $\gamma$  is the angle value in the *RCSZE* coordinate. Then

$$\alpha = (\alpha - \beta) + (\beta - \gamma) + \gamma. \tag{1}$$

Herein,  $(\alpha - \beta)$  is the difference between the mechanical zero of system and the absolute zero of encoder, it is a constant once the encoder is fixed on the mechanical system;  $(\beta - \gamma)$  is the angle difference between the *RCSZE* and the *RCAZE*;  $\gamma$  is the

instant position value directly acquired by the DAQ card. In fact,  $(\beta-\gamma)$  is also a constant once the system is switching-on, which will keep unchanged until the system is switching-on again. According to Eq.(1), therefore,  $\alpha$  can be calculated if  $(\beta-\gamma)$  could be obtained.

In order to get the value of  $(\beta-\gamma)$ , two hardware channels are required to acquire the same signal of the encoder (see Fig.5). The reference pulse sent from *Channel 1* is ignored, but not for the reference pulse sent from *Channel 2*. When the reference pulse occurs, the counter of *Channel 2* is cleared. Based on the different setting for *Channel 1* and *Channel 2*, a novel zero-finding algorithm is proposed.



**Fig. 5.** Schematic block of the zero-finding algorithm. Where, the *WoR* and *WR* are two self-defined input variables of the *FindingZero* Block which is developed by S-function. Because a single pulse is equal to five-ten thousandths degree of angle, thus, the value *K* of *Gain1* and *Gain2* is 0.00005, which is used to convert the number of pulse into the angle value. The constant *C* is the difference between the mechanical zero of system and the absolute zero of encoder.

In Fig.5, *Channel 1* and *Channel 2* are two independent data acquisition channels of PCI-6601. When the encoder is revolving clockwise, the value of *WoR* is increasing monotonously, and the value of *WR* is also increasing monotonously except the moment when the output is cleared to zero at the position of reference marks; When the encoder is revolving anti-clockwise, the value of *WoR* is decreasing monotonously, and the value of *WR* is also decreasing monotonously except the moment when the output is cleared to zero at the position of reference marks. Therefore, it is concluded that a reference mark is traversing across when the absolute value of *WR* is decreasing. At this time, the value  $\gamma_k$  of this reference mark can be obtained by subtracting *WR* from *WoR*.

$$\gamma_k = WoR_k - WR_k \tag{2}$$

In the same way, the value  $\gamma_{k+1}$  of the next reference mark can be obtained.

$$\gamma_{k+1} = WoR_{k+1} - WR_{k+1} \tag{3}$$

Therefore, the angle difference between the reference mark and its neighbor reference mark is obtained.

$$\Delta_{k,k+1} = \gamma_{k+1} - \gamma_k. \tag{4}$$

Thus, the value of  $(\beta - \gamma)$  can be worked out according to  $\Delta_{k,k+1}$  and Fig.2. Withal,

$$\alpha - \gamma = (\alpha - \beta) + (\beta - \gamma). \tag{5}$$

where,  $(\alpha - \gamma)$  is the angle difference between the *RCAZE* and the *RCMZS*. Therefore, the system will arrive at the mechanical zero position by rotating  $-(\alpha - \gamma)^\circ$  from the start zero position. Fig.6 is the flowchart of the zero-finding algorithm.

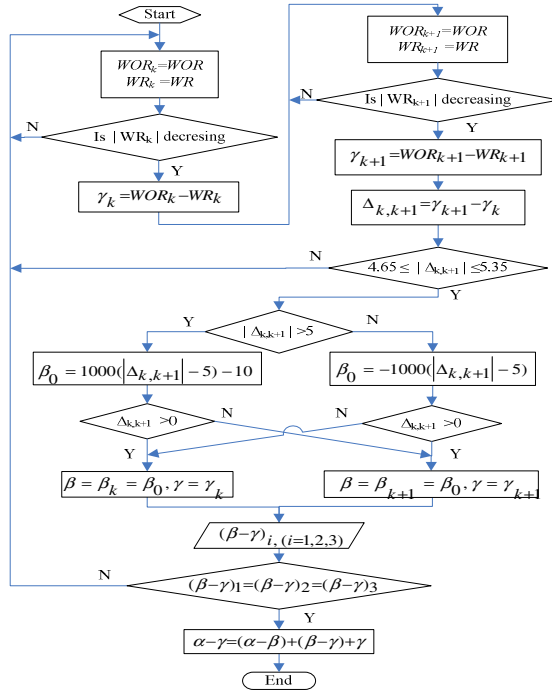
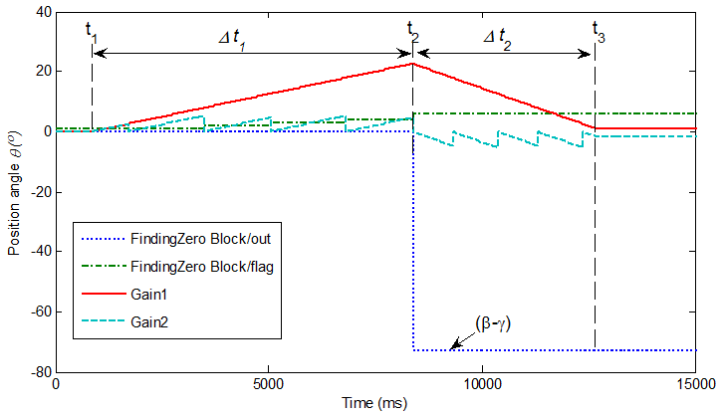


Fig. 6. Flowchart of the zero-finding algorithm

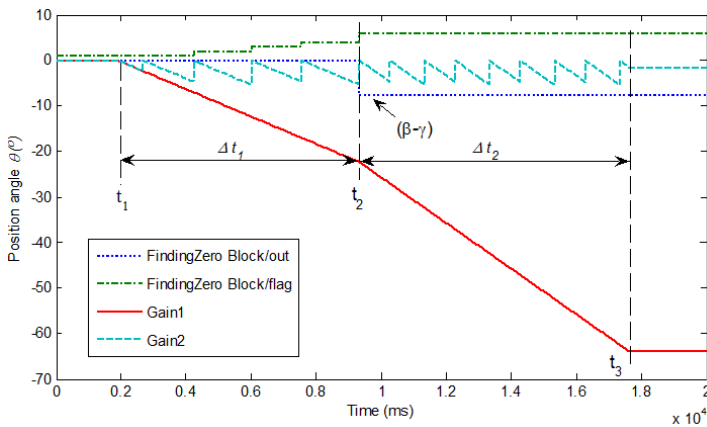
**Illustration:** if  $\Delta_{k,k+1}$  is equal to 503 lines, it can be concluded that the angle range traversed must be  $20^\circ \sim 25.03^\circ$ , and the instant position must be one of both limits ( $20^\circ$  or  $25.03^\circ$ ). Keeping the encoder rotating on, we can obtain the position value of the neighbor reference mark. The difference between them must be 497 lines or 498 lines (no other choice). If the difference is 497 lines, the instant position value of this moment is deduced to be  $30^\circ$ ; otherwise, the instant position value of this moment is  $15.02^\circ$ . The direction of motion can also be judged according to the change trend of the  $\Delta$ . Furthermore, if an error occurs during the zero-finding process, all need to do is just that the encoder is kept rotating on for  $5^\circ$ . The anti-interference ability of the zero-finding is improved greatly.

## 5 Experimental Results and Analysis

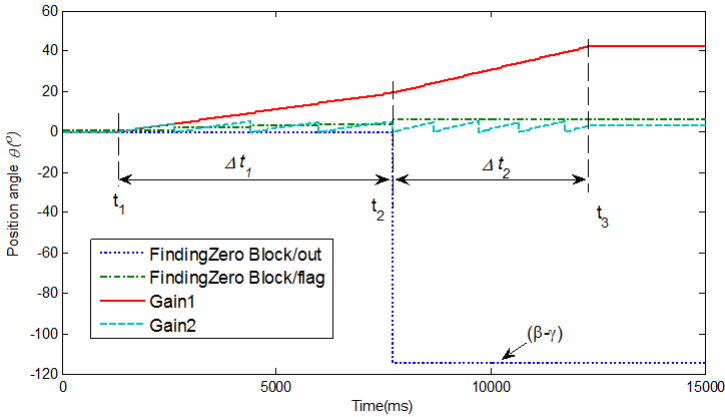
Several experiments are carried out on the outer frame of the three-axis turntable at some special positions. The experimental results are shown in Fig.7-9. It is indicated that the zero-finding process mainly includes two processes: ①  $\Delta t_1$  - finding the correction value  $(\beta-\gamma)$ , ②  $\Delta t_2$  - reaching to the mechanical zero position. The system commences to seek the correction value at the  $t_1$  moment, and keeps swinging to find the sequential  $(\beta-\gamma)$  at the speed of  $3^\circ/\text{s}$ . If the values of  $(\beta-\gamma)$  obtained in three times are equal, the correction value is considered to be available. Then, the system steps into the next process at the  $t_2$  moment, in which the system reaches to the mechanical zero position at the speed of  $5^\circ/\text{s}$ . The mechanical zero position arrived at the  $t_3$  moment.



**Fig. 7.** Experimental result for zero-finding at about zero position. The solid line denotes the trajectory of the encoder. The correction value  $(\beta-\gamma)$  is ascertained at the  $t_2$  moment, which only needs to rotate about  $20^\circ$ . The encoder reaches to mechanical zero position during  $\Delta t_2$ .



**Fig. 8.** Experimental result for zero-finding at position infinite position



**Fig. 9.** Experimental result for zero-finding at about negative infinite position

The meanings of signals in Fig.7-9 are the same as described in Fig.5. The output of *Gain1* and *Gain2* is corresponding to the angle value acquired by *Channel 1* and *Channel 2*, respectively. The solid line (*Gain1*) is the output of *Gain1* block in which the reference pulse is ignored. The dash line (*Gain2*) is the output of *Gain2* block in which the reference pulse is not ignored, which will be cleared to zero when a reference pulse occurs. The dotted line is the out of FindingZero Block, which denotes the angle difference between the *RCSZE* (Reference Coordinates of Start Zero of Encode) and the *RCAZE* (Reference Coordinates of Absolute Zero of Encode), namely  $(\beta-\gamma)$ . The dash-dotted line is the flag if the correction value  $(\beta-\gamma)$  is found.

## 6 Conclusion

Addressing a practical problem, a novel zero-finding algorithm is proposed based on the special coding mechanism. The experimental results and analysis show that the proposed algorithm have not only greatly speeded up and simplified the operation of zero-finding, but also improved the accuracy, the robustness and the ability to anti-interference. After the zero-finding, the incremental encoder with “memory ability” can supply the absolute position value as the absolute encoder could do. This paper provides another effective method of position orientating by the incremental angle encoder, which has important value to practical applications for engineers. In addition, the proposed algorithm can be adopted in any related case, not limited for hydraulic servo systems.

**Acknowledgments.** We are grateful for the help offered by X. F. ZHA, a Prof. from National Institute of Standards and Technology (NIST), Maryland University, USA. This work was supported by a grant from the National Natural Science Foundation of China (No.50975172).

## References

1. Liu, W.K., Shi, J.L.: The Application of Optical Rotary Encoder in Angle Measure. *Modern Manufacturing Engineering* 11, 90–93 (2006) (in Chinese)
2. Marco, F., Paolo, G., Francesco, P., Roberto, P., Marco, T.: An Embedded System for Position and Speed Measurement Adopting Incremental Encoders. In: *Conference Record of the IEEE Industry Applications Conference, 39th IAS Annual Meeting*, pp. 1192–1199 (2004)
3. Sugimoto, H., Ichikawa, T.: Pole Sensorless Control for Brushless DC Servomotor with Incremental Encoder. In: *The 4th Transactions of the 1999 National Convention Record*, pp. 482–483. IEE Press, Japan (1999)
4. Sugimoto, H., Ichikawa, T., Hosoi, K., Kawasaki, S.: Magnetic Pole Position Detection Method and Control of a Brushless DC Servomotor with Incremental Encoder. *Electrical Engineering in Japan* 145, 64–77 (2003)
5. Xie, Y., Tseng, K.J.: An Observer-Based Robust Adaptive Controller for Permanent Magnet Synchronous Motor Drive with Initial Rotor Angle Uncertainty. *IEEE Transactions on Energy Conversion* 20(1), 11–120 (2005)
6. Xu, Z., Hu, X.D., Luo, C.Z., Gu, L.: The Electronic Utliplication and Interpolation of Incremental Optical Encoder with Reference Mark. *Acta Photonica Sinica* 31(12), 1497–1500 (2002)
7. Gao, Z.Q., Li, B.F.: Examples of Revolving Axis Positioning with Incremental Encoder. *Mechanical Engineer* 8, 133–134 (2007) (in Chinese)
8. Brochure, H.: Angle Encoders with Integral Bearing,  
[http://www.heidenhain.com/index.php?WCMSGGroup\\_2238\\_177=2243&WCMSGGroup\\_2243\\_177=846591\\_109\\_21\\_18458.pdf](http://www.heidenhain.com/index.php?WCMSGGroup_2238_177=2243&WCMSGGroup_2243_177=846591_109_21_18458.pdf)

# Decentralized Synchronization and Output Tracking Control of Nondiffusively Coupled Complex Dynamical Networks

Gequn Liu<sup>1,2</sup>, Xiaoming Xu<sup>1,2,3</sup>, and Lei Liu<sup>1</sup>

<sup>1</sup> School of Management, University of Shanghai for Science and Technology, Shanghai, China

<sup>2</sup> Shanghai Academy of Systems Science, Shanghai, China

<sup>3</sup> Department of Automation, Shanghai Jiao Tong University, Shanghai, China  
lgqpapers@sina.com

**Abstract.** Diffusive coupling configuration of complex dynamical networks derives from the situation that nodes are coupled with states difference between each other. For the limitation of applicability of diffusive coupling model, it is necessary to study the control problem of nondiffusively coupled complex networks. A decentralized synchronization criterion with state feedback control scheme was proposed based on linear matrix inequality methodology. A simple criterion for the verification of decentralized stabilizability of the network is given. Furthermore, a decentralized output tracking control method is proposed based on the former synchronization criterion. Finally a nondiffusively coupled scale-free network is provided as the example to verify the effectiveness of the given methods.

**Keywords:** complex networks, nondiffusive coupling, synchronization control, output tracking control, linear matrix inequality.

## 1 Introduction

Synchronization of complex dynamic networks<sup>[1-8]</sup> has attracted widely attentions for its scientific importance and application values. The most investigated network models have diffusive coupling<sup>[4-8]</sup>. Diffusive coupling means that the row sums of outer coupling matrix are zeros. Diffusive coupling configuration derives from the situation that nodes are coupled with states difference between each other. Zhisheng Duan pointed that subsystems in diffusive coupled networks may vary with different coupling structures. The reason is that dynamics of subsystem is no longer  $\dot{\mathbf{x}}_i = f(\mathbf{x}_i)$

but  $\dot{\mathbf{x}}_i = f(\mathbf{x}_i) + ca_{ii}\Gamma\mathbf{x}_i$ , where  $a_{ii} = -\sum_{j=1, j \neq i}^N a_{ij}$  is changed with interconnections<sup>[8]</sup>. In

most cases, node dynamics remain unchanged with interconnections for they are coupled not by difference between states but by states directly. For example, in product price networks, the price of product A is countered as cost of product B directly. So it is of application meaning to study the control of networks with nondiffusive coupling.



In [4-7], synchronization of diffusive coupled networks was studied via linearization method named as “master stability function”. Paper [8] proposed a synchronization method based on state feedback control scheme. Decentralized stabilization algorithms were given for large scale systems in [8-13], decentralized tracking control algorithms were investigated for large scale systems in [12-14]. By contrast, there is little research on output tracking control of complex networks so far. In view of the importance of tracking control problem, decentralized synchronization and output tracking control of nondiffusively coupled complex networks are going to be investigated in this paper.

The rest of the paper is outlined as follows. A nondiffusively coupled complex networks model is introduced with some assumptions and preliminaries in Section 2. In section 3, decentralized synchronization of the network is investigated and as a result two useful criteria were given. A condition for decentralized output tracking control of the network is proposed in section 4. In section 5 an example is provided to verify the effectiveness of the methods proposed. Section 6 is the conclusion.

## 2 Problem Formulation and Some Preliminaries

Consider a nondiffusively coupled complex dynamical network with  $N$  identical continuous time invariant linear nodes

$$\begin{cases} \dot{\mathbf{x}}_i = \mathbf{A}_0 \mathbf{x}_i + \mathbf{B}_0 \mathbf{u}_i + \sum_{j=1, j \neq i}^N h_{ij} \Gamma \mathbf{x}_j \\ \mathbf{y}_i = \mathbf{C}_0 \mathbf{x}_i \quad i = 1, 2, \dots, N \end{cases}, \quad (1)$$

where  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})^T \in \mathbf{R}^n$ ,  $\mathbf{u}_i = (u_{i1}, u_{i2}, \dots, u_{im})^T \in \mathbf{R}^m$ ,  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{ir})^T \in \mathbf{R}^r$ ,  $\mathbf{A}_0 \in \mathbf{R}^{n \times n}$ ,  $\mathbf{B}_0 \in \mathbf{R}^{n \times m}$ ,  $\mathbf{C}_0 \in \mathbf{R}^{r \times n}$  are state vector, control input, output, system matrix, input matrix and output matrix of  $i$  th node respectively.  $\Gamma = (\gamma_{ij})_{n \times n} \in \mathbf{R}^{n \times n}$  is the inner coupling matrix. Outer coupling matrix  $\mathbf{H} = (h_{ij})_{N \times N} \in \mathbf{R}^{N \times N}$  denotes the coupling configuration of the entire network. If there is a link between node  $i$  and node  $j$  ( $i \neq j$ ),  $h_{ij} = h_{ji} > 0$  is the coupling strength; otherwise,  $h_{ij} = h_{ji} = 0$ . The diagonal element of  $\mathbf{H}$  is defined as  $h_{ii} = 0$ .

Suppose that the network is connected in the sense of with no isolated clusters. As a result  $\mathbf{H}$  is a real symmetric irreducible matrix. The difference between network (1) defined here and diffusively coupled network is that the outer coupling matrix in (1) has zero diagonal elements, beside, the row sums of  $\mathbf{H}$  are not equal to 0.

The state equation of  $i$  th isolated node is

$$\begin{cases} \dot{\mathbf{x}}_i = \mathbf{A}_0 \mathbf{x}_i + \mathbf{B}_0 \mathbf{u}_i \\ \mathbf{y}_i = \mathbf{C}_0 \mathbf{x}_i \end{cases}. \quad (2)$$

Following assumptions<sup>[9-12]</sup> for system (2) are necessary:

*Assumption 1.*  $(\mathbf{A}_0, \mathbf{B}_0)$  is controllable and  $(\mathbf{A}_0, \mathbf{C}_0)$  is detectable.

*Assumption 2.*  $\text{rank} \begin{bmatrix} A_0 & B_0 \\ C_0 & 0 \end{bmatrix} = n + r$ .

The objective of decentralized synchronization is to utilize state feedback control law

$$u_i = Kx_i \tag{3}$$

to stabilize network (1) onto the equilibrium point  $x_1 = x_2 = \dots = x_N = 0$  when  $t \rightarrow \infty$ . Where  $K \in \mathbf{R}^{n \times n}$  is the state feedback matrix of  $i$  th node.

The objective of decentralized output tracking control is to utilize control law

$$u_i = f(x_i, y_i, y_R) \tag{4}$$

to drive output  $y_i (i = 1, 2, \dots, N)$  asymptotically tacking any given constant reference signal  $y_R \in \mathbf{R}^r$ .

Two lemmas<sup>[15-16]</sup> are given firstly for following application.

*Lemma 1.* (Schur complement) Suppose that a symmetric matrix is partitioned as  $S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$ , where  $S_{11}$  and  $S_{22}$  are square. Then following conditions are equivalent to each other:

- (1)  $S < 0$ ;
- (2)  $S_{11} < 0, S_{22} - S_{12}^T S_{11}^{-1} S_{12} < 0$ ;
- (3)  $S_{22} < 0, S_{11} - S_{12}^T S_{22}^{-1} S_{12} < 0$ .

*Lemma 2.* For any vectors  $x \in \mathbf{R}^n, y \in \mathbf{R}^m$  and positive definite matrix  $M \in \mathbf{R}^{n \times n}$ , the following LMI holds for any positive definite matrix  $G \in \mathbf{R}^{m \times m}$

$$2x^T M y \leq x^T M G^{-1} M^T x + y^T G y. \tag{5}$$

### 3 Decentralized Synchronization

Rewrite close loop feedback control system (1),(3) in compact form :

$$\dot{x} = [I_N \otimes (A_0 + B_0 K) + H \otimes \Gamma] x = A_c x, \tag{6}$$

where  $\otimes$  is the Kronecker product,  $x = [x_1^T, x_2^T, \dots, x_N^T]^T$ ,

$$A_c = \begin{bmatrix} A_0 + B_0 K & h_{12} \Gamma & \dots & h_{1N} \Gamma \\ h_{21} \Gamma & A_0 + B_0 K & \dots & h_{2N} \Gamma \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1} \Gamma & h_{N2} \Gamma & \dots & A_0 + B_0 K \end{bmatrix}. \tag{7}$$

*Theorem 1.* For complex network (1), if there are matrix  $S \in \mathbf{R}^{m \times n}$  and positive definite matrix  $Q = Q^T \in \mathbf{R}^{m \times n}$ , such that for  $i = 1, 2, \dots, N$

$$\begin{bmatrix} \Theta_i & Q \\ Q & \frac{-1}{N-1} I_n \end{bmatrix} < 0, \tag{8}$$

$$\text{where } \Theta_i = \mathbf{Q}\mathbf{A}_0^T + \mathbf{A}_0\mathbf{Q} + \mathbf{S}^T\mathbf{B}_0^T + \mathbf{B}_0\mathbf{S} + \mathbf{\Gamma}\mathbf{\Gamma}^T \sum_{j=1, j \neq i}^N h_{ij}^2, \quad (9)$$

network (1) is stabilizable under decentralized state feedback control, the close loop control system (6) is asymptotically stable with control law

$$\mathbf{u}_i = \mathbf{K}\mathbf{x}_i = \mathbf{S}\mathbf{Q}^{-1}\mathbf{x}_i \quad i = 1, 2, \dots, N. \quad (10)$$

*Proof.* Construct Lyapunov function

$$\mathbf{V}(\mathbf{x}, t) = \mathbf{x}^T \mathbf{P}\mathbf{x}, \quad (11)$$

where  $\mathbf{P} = \text{diag}(\mathbf{P}_1, \mathbf{P}_1, \dots, \mathbf{P}_1) \in \mathbf{R}^{n \times nN}$  is positive definite matrix,  $\mathbf{P}_1 = \mathbf{P}_1^T \in \mathbf{R}^{n \times n}$ ,  $i = 1, 2, \dots, N$ .

Obviously, (11) is positive definite.

The time derivation of  $\mathbf{V}(\mathbf{x}, t)$  along system (6) is

$$\begin{aligned} \dot{\mathbf{V}}(\mathbf{x}, t) &= \dot{\mathbf{x}}^T \mathbf{P}\mathbf{x} + \mathbf{x}^T \mathbf{P}\dot{\mathbf{x}} \\ &= \mathbf{x}^T \left[ [\mathbf{I}_N \otimes (\mathbf{A}_0 + \mathbf{B}_0\mathbf{K})]^T \mathbf{P} + \mathbf{P}[\mathbf{I}_N \otimes (\mathbf{A}_0 + \mathbf{B}_0\mathbf{K})] \right] \mathbf{x} + 2\mathbf{x}^T \mathbf{P}(\mathbf{H} \otimes \mathbf{\Gamma})\mathbf{x} \\ &= \sum_{i=1}^N \mathbf{x}_i^T \left[ (\mathbf{A}_0 + \mathbf{B}_0\mathbf{K})^T \mathbf{P}_1 + \mathbf{P}_1(\mathbf{A}_0 + \mathbf{B}_0\mathbf{K}) \right] \mathbf{x}_i + \sum_{i=1}^N \sum_{j=1, j \neq i}^N 2\mathbf{x}_i^T \mathbf{P}_1 h_{ij} \mathbf{\Gamma}\mathbf{x}_j \end{aligned} \quad (12)$$

Similar to [13], from lemma 2 we have

$$2\mathbf{x}_i^T \mathbf{P}_1 h_{ij} \mathbf{\Gamma}\mathbf{x}_j \leq \mathbf{x}_i^T \mathbf{P}_1 h_{ij} \mathbf{\Gamma}\mathbf{\Gamma}^T h_{ij} \mathbf{P}_1 \mathbf{x}_i + \mathbf{x}_j^T \mathbf{I}_n \mathbf{x}_j = \mathbf{x}_i^T \mathbf{P}_1 h_{ij}^2 \mathbf{\Gamma}\mathbf{\Gamma}^T \mathbf{P}_1 \mathbf{x}_i + \mathbf{x}_j^T \mathbf{I}_n \mathbf{x}_j. \quad (13)$$

Substitute (13) into (12) we have

$$\begin{aligned} \dot{\mathbf{V}}(\mathbf{x}, t) &\leq \sum_{i=1}^N \mathbf{x}_i^T \left[ (\mathbf{A}_0 + \mathbf{B}_0\mathbf{K})^T \mathbf{P}_1 + \mathbf{P}_1(\mathbf{A}_0 + \mathbf{B}_0\mathbf{K}) \right] \mathbf{x}_i \\ &\quad + \sum_{i=1}^N \mathbf{x}_i^T \mathbf{P}_1 \left( \sum_{j=1, j \neq i}^N h_{ij}^2 \right) \mathbf{\Gamma}\mathbf{\Gamma}^T \mathbf{P}_1 \mathbf{x}_i + (N-1) \sum_{i=1}^N \mathbf{x}_i^T \mathbf{x}_i \\ &= \sum_{i=1}^N \mathbf{x}_i^T \mathbf{P}_1 \left[ \mathbf{P}_1^{-1}(\mathbf{A}_0 + \mathbf{B}_0\mathbf{K})^T + (\mathbf{A}_0 + \mathbf{B}_0\mathbf{K})\mathbf{P}_1^{-1} + \mathbf{\Gamma}\mathbf{\Gamma}^T \sum_{j=1, j \neq i}^N h_{ij}^2 + (N-1)\mathbf{P}_1^{-1}\mathbf{P}_1^{-1} \right] \mathbf{P}_1 \mathbf{x}_i \end{aligned} \quad (14)$$

According to Lyapunov stability theory, for  $i = 1, 2, \dots, N$ , system (6) is asymptotically stable at equilibrium point 0 when following LMIs holds

$$\mathbf{P}_1^{-1}(\mathbf{A}_0 + \mathbf{B}_0\mathbf{K})^T + (\mathbf{A}_0 + \mathbf{B}_0\mathbf{K})\mathbf{P}_1^{-1} + \mathbf{\Gamma}\mathbf{\Gamma}^T \sum_{j=1, j \neq i}^N h_{ij}^2 + (N-1)\mathbf{P}_1^{-1}\mathbf{P}_1^{-1} < 0. \quad (15)$$

Denote

$$\mathbf{Q} = \mathbf{P}_1^{-1}, \quad (16)$$

$$\mathbf{S} = \mathbf{K}\mathbf{P}_1^{-1}, \quad (17)$$

equation (15) can be reformatted as

$$\mathbf{Q}\mathbf{A}_0^T + \mathbf{A}_0\mathbf{Q} + \mathbf{S}^T\mathbf{B}_0^T + \mathbf{B}_0\mathbf{S} + \mathbf{\Gamma}\mathbf{\Gamma}^T \sum_{j=1, j \neq i}^N h_{ij}^2 + (N-1)\mathbf{Q}\mathbf{I}_n\mathbf{Q} < 0. \quad (18)$$

From lemma 1, LMI (18) is equivalent to LMI (8).

From (16) and (17), we have  $\mathbf{K} = \mathbf{S}\mathbf{Q}^{-1}$ . □

It can be easily find that the only deference between  $N$  lower dimensional LMIs in (8), (9) is the sum of coupling strength square  $\sum_{j=1, j \neq i}^N h_{ij}^2$  of  $i$  th node. So a question appears, can we find a feasible region of  $\sum_{j=1, j \neq i}^N h_{ij}^2$  for (8) to simplify the verification of decentralized stability of network (1).

*Theorem 2.* For complex network (1), if there exist optimal solution  $\mathbf{S}_{\max} \in \mathbf{R}^{n \times n}$ ,  $\mathbf{Q}_{\max} \in \mathbf{R}^{n \times n}$  and  $h_{\max} \in \mathbf{R}$  for

$$\begin{aligned} & \min \quad -h \\ & \text{s.t.} \quad \begin{cases} \begin{bmatrix} \boldsymbol{\Theta} & \mathbf{Q} \\ \mathbf{Q} & \frac{-1}{N-1} \mathbf{I}_n \end{bmatrix} < 0, \\ \mathbf{Q} = \mathbf{Q}^T > 0, \quad h > 0 \end{cases} \end{aligned} \quad (19)$$

and if there exist optimal solution  $\mathbf{S}_{\min} \in \mathbf{R}^{n \times n}$ ,  $\mathbf{Q}_{\min} \in \mathbf{R}^{n \times n}$  and  $h_{\min} \in \mathbf{R}$  for

$$\begin{aligned} & \min \quad h \\ & \text{s.t.} \quad \begin{cases} \begin{bmatrix} \boldsymbol{\Theta} & \mathbf{Q} \\ \mathbf{Q} & \frac{-1}{N-1} \mathbf{I}_n \end{bmatrix} < 0, \\ \mathbf{Q} = \mathbf{Q}^T > 0, \quad h > 0 \end{cases} \end{aligned} \quad (20)$$

in (19) and (20)

$$\boldsymbol{\Theta} = \mathbf{Q}\mathbf{A}_0^T + \mathbf{A}_0\mathbf{Q} + \mathbf{S}^T\mathbf{B}_0^T + \mathbf{B}_0\mathbf{S} + \Gamma\Gamma^T h. \quad (21)$$

For  $i = 1, 2, \dots, N$ , if following inequation holds

$$h_{\min} \leq \sum_{j=1, j \neq i}^N h_{ij}^2 \leq h_{\max}, \quad (22)$$

network (1) is stabilizable under decentralized state feedback control.

The proof is omitted for (19) and (20) are convex optimization<sup>[16-17]</sup> problems.

## 4 Decentralized Output Tracking Control

Similar to [14], construct augmented state vector

$$\mathbf{z}_i = \begin{bmatrix} \mathbf{x}_i \\ \mathbf{q}_i \end{bmatrix} \quad i = 1, 2, \dots, N, \quad (23)$$

where

$$\dot{\mathbf{q}}_i = \mathbf{y}_i - \mathbf{y}_R = \mathbf{C}_0 \mathbf{x}_i - \mathbf{y}_R. \quad (24)$$

Then construct matrix

$$\bar{\mathbf{A}}_0 = \begin{bmatrix} \mathbf{A}_0 & 0 \\ \mathbf{C}_0 & 0 \end{bmatrix}, \bar{\mathbf{B}}_0 = \begin{bmatrix} \mathbf{B}_0 \\ 0 \end{bmatrix}, \bar{\mathbf{F}} = \begin{bmatrix} \mathbf{F} & 0 \\ 0 & 0 \end{bmatrix}, \bar{\mathbf{w}} = \begin{bmatrix} 0 \\ \mathbf{y}_R \end{bmatrix}, \quad (25)$$

we have a new network with augmented nodes

$$\dot{\mathbf{z}}_i = \bar{\mathbf{A}}_0 \mathbf{z}_i + \bar{\mathbf{B}}_0 \mathbf{u}_i + \sum_{j=1, j \neq i}^N h_{ij} \bar{\mathbf{F}} \mathbf{z}_j + \bar{\mathbf{w}} \quad i=1, 2, \dots, N. \quad (26)$$

*Theorem 3.* For complex network (1), if there are matrix  $\mathbf{S} \in \mathbf{R}^{m \times (n+r)}$  and positive definite matrix  $\mathbf{Q} = \mathbf{Q}^T \in \mathbf{R}^{(n+r) \times (n+r)}$ , such that for  $i=1, 2, \dots, N$

$$\begin{bmatrix} \boldsymbol{\theta}_i & \mathbf{Q} \\ \mathbf{Q} & \frac{-1}{N-1} \mathbf{I}_{n+r} \end{bmatrix} < 0, \quad (27)$$

where 
$$\boldsymbol{\theta}_i = \mathbf{Q} \bar{\mathbf{A}}_0^T + \bar{\mathbf{A}}_0 \mathbf{Q} + \mathbf{S}^T \bar{\mathbf{B}}_0^T + \bar{\mathbf{B}}_0 \mathbf{S} + \bar{\mathbf{F}} \bar{\mathbf{F}}^T \sum_{j=1, j \neq i}^N h_{ij}^2, \quad (28)$$

the output of network (1) can asymptotically track reference signal  $\mathbf{y}_R$  under the decentralized control law

$$\mathbf{u}_i = \mathbf{K} \mathbf{z}_i = \mathbf{S} \mathbf{Q}^{-1} \mathbf{z}_i \quad i=1, 2, \dots, N. \quad (29)$$

*Proof.* Rewriting close loop feedback control system (26), (29) in compact form

$$\dot{\mathbf{z}} = [\mathbf{I}_N \otimes (\bar{\mathbf{A}}_0 + \bar{\mathbf{B}}_0 \mathbf{K}) + \mathbf{H} \otimes \bar{\mathbf{F}}] \mathbf{z} + \mathbf{I}_N \otimes \bar{\mathbf{w}} = \bar{\mathbf{A}}_C \mathbf{z} + \mathbf{I}_N \otimes \bar{\mathbf{w}}, \quad (30)$$

where  $\mathbf{z} = [\mathbf{z}_1^T, \mathbf{z}_2^T, \dots, \mathbf{z}_N^T]^T$ ,

$$\bar{\mathbf{A}}_C = \begin{bmatrix} \bar{\mathbf{A}}_0 + \bar{\mathbf{B}}_0 \mathbf{K} & h_{12} \bar{\mathbf{F}} & \dots & h_{1N} \bar{\mathbf{F}} \\ h_{21} \bar{\mathbf{F}} & \bar{\mathbf{A}}_0 + \bar{\mathbf{B}}_0 \mathbf{K} & \dots & h_{2N} \bar{\mathbf{F}} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1} \bar{\mathbf{F}} & h_{N2} \bar{\mathbf{F}} & \dots & \bar{\mathbf{A}}_0 + \bar{\mathbf{B}}_0 \mathbf{K} \end{bmatrix}. \quad (31)$$

Differentiating both sides of (30), we have

$$\dot{\mathbf{z}} = \bar{\mathbf{A}}_C \mathbf{z}. \quad (32)$$

From assumption 2 we know that system (26) is totally controllable. For system (32) has the same form as system (6), under the condition of theorem 3, from theorem 1 we know that system (32) is asymptotically stable at equilibrium point 0, which means

$$\lim_{t \rightarrow \infty} (\dot{\mathbf{z}}_i - 0) = \lim_{t \rightarrow \infty} \left( \begin{bmatrix} \dot{\mathbf{x}}_i \\ \dot{\mathbf{q}}_i \end{bmatrix} - 0 \right) = 0, \quad (33)$$

namely 
$$\lim_{t \rightarrow \infty} \dot{\mathbf{q}}_i = \lim_{t \rightarrow \infty} (\mathbf{y}_i - \mathbf{y}_R) = 0. \quad (34)$$

□

### 5 Example

Consider isolate linear time invariant node with state equation

$$\begin{cases} \dot{\mathbf{x}}_i = \begin{bmatrix} -7 & 2 \\ 1 & -6 \end{bmatrix} \mathbf{x}_i + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \mathbf{u}_i \\ \mathbf{y}_i = [1 \ 1] \mathbf{x}_i \end{cases} \quad (35)$$

Let original nodes number and newly added edges number be 2, generate a nondiffusively coupled scale-free network of  $N = 10$  nodes by B-A model<sup>[1]</sup>. The configuration of the network is given in Figure 1.

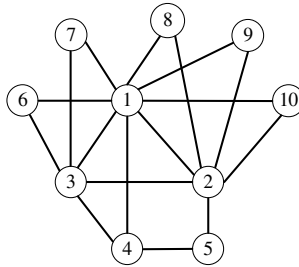


Fig. 1. Configuration of the scale-free network

The outer coupling matrix is

$$H = \begin{bmatrix} 0 & 1 & 1 & 2 & 0 & 1 & 1 & 2 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 2 \\ 1 & 1 & 0 & 2 & 0 & 1 & 1 & 0 & 0 & 0 \\ 2 & 0 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (36)$$

the inner coupling matrix is  $\Gamma = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ , initial state of the network is

$$[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] = \begin{bmatrix} 1 & 2 & -4 & -1 & 1 & 0 & 2 & -7 & -1 & 0 \\ -1 & 0 & -1 & 2 & 3 & -2 & -3 & 2 & 2 & 4 \end{bmatrix}$$

### 5.1 Decentralized Stability Checking

Solving the optimization problem (19) and (20), we have

$$\begin{cases} \mathbf{S}_{\max h} = \begin{bmatrix} -20195 & -20188 \end{bmatrix} \\ \mathbf{Q}_{\max h} = \begin{bmatrix} 38.4395 & 37.5506 \\ 37.5506 & 38.4395 \end{bmatrix} \\ h_{\max} = 14.2222 \end{cases} \text{ and } \begin{cases} \mathbf{S}_{\min h} = \begin{bmatrix} -1.2449 \times 10^8 & -1.2449 \times 10^8 \end{bmatrix} \\ \mathbf{Q}_{\min h} = \begin{bmatrix} 2143.7 & 2142.5 \\ 2142.5 & 2143.7 \end{bmatrix} \\ h_{\min} = 1.3818 \times 10^{-12} \end{cases} .$$

It is easy to know that the coupling strength of the network satisfies  $h_{\min} < 2 \leq \sum_{j=1, j \neq i}^N h_{ij}^2 \leq 14 < h_{\max}$ , from theorem 2, the network is stabilizable under decentralized state feedback control scheme.

### 5.2 Decentralized Synchronization

Solve LMI (8) we have

$$\mathbf{Q} = \begin{bmatrix} 1.3717 & 0.4939 \\ 0.4939 & 1.3726 \end{bmatrix}, \mathbf{S} = [-170.294 \quad -163.544], \mathbf{K} = [-93.34 \quad -85.56].$$

The state curves of the decentralized synchronized network are given in figure 2.

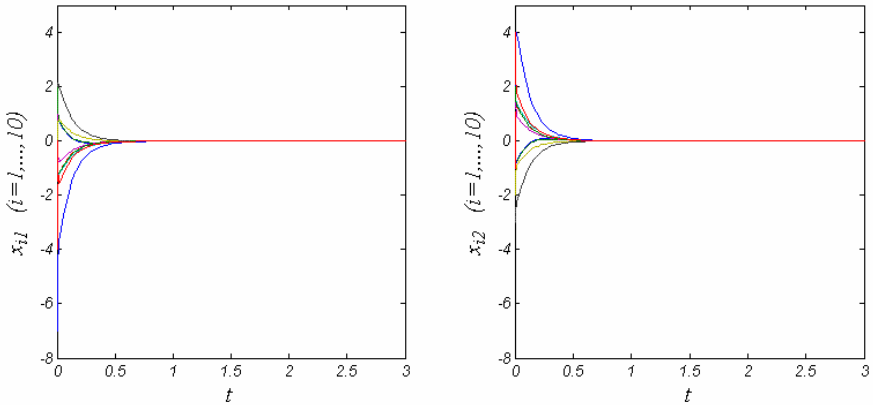


Fig. 2. State curves of the decentralized synchronized network

### 5.3 Decentralized Output Tracking

Consider constant reference signal  $\mathbf{y}_R = 2$ , we have

$$\text{rank} \begin{bmatrix} \mathbf{A}_0 & \mathbf{B}_0 \\ \mathbf{C}_0 & 0 \end{bmatrix} = \text{rank} \begin{bmatrix} -7 & 2 & 1 \\ 1 & -6 & 1 \\ 1 & 1 & 0 \end{bmatrix} = n + r = 3,$$

$$\bar{A}_0 = \begin{bmatrix} -7 & 2 & 0 \\ 1 & -6 & 0 \\ 1 & 1 & 0 \end{bmatrix}, \bar{B}_0 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \bar{F} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \bar{w} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}.$$

solve LMI (27), we have

$$Q = \begin{bmatrix} 1.3075 & 0.4146 & -0.1055 \\ 0.4146 & 1.3069 & -0.1055 \\ -0.1055 & -0.1055 & 0.0410 \end{bmatrix},$$

$$S = [-100.7637 \quad -96.4507 \quad -0.5674],$$

$$K = [-87.2193 \quad -82.4767 \quad -450.143].$$

The decentralized tracking controller is

$$u_i = -87.2193x_{i1} - 82.4767x_{i2} - 450.143 \int_0^t (y_i(\tau) - y_R) d\tau.$$

The output curve and control input curve of the close loop network are given in figure 3.

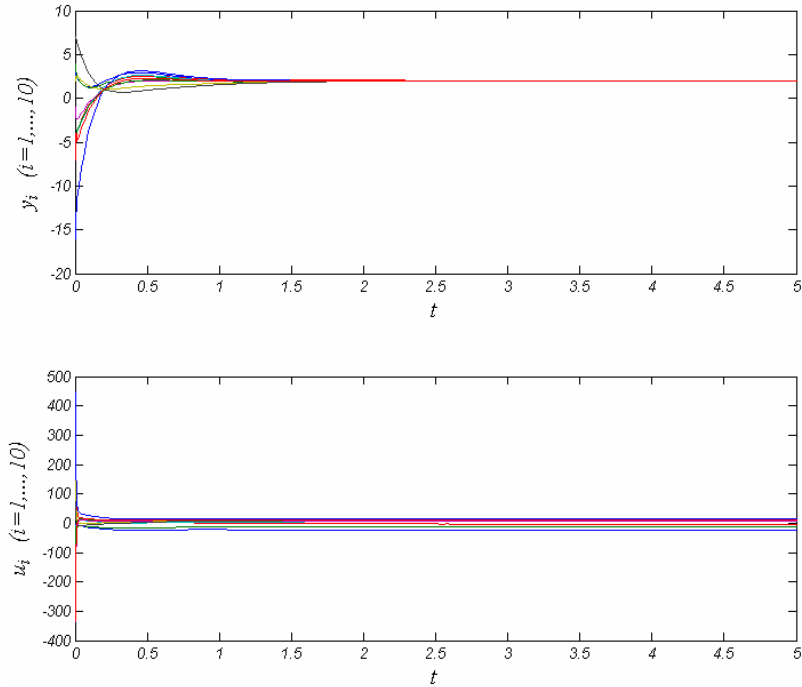


Fig. 3. Output curve and control input curve of the close loop network

## 6 Conclusions

Synchronization and output tracking control of nondiffusively coupled complex networks are investigated in this paper. Two general criterions are given to ensure the asymptotically stability of close loop control system. A simple criterion with



optimization problem form is given for the purpose of quick verification of decentralized stabilizability of the network. The network model studied here is of nondiffusive coupling topology, which makes the obtained results have a wider application area than those of diffusively coupled networks. The tracking control method developed can serve as a start point of manipulation technique of complex networks.

**Acknowledgments.** This work was partly supported by Shanghai Key Scientific and Technological Project (072305126), Shanghai Leading Discipline Project (S30501), Shanghai Basic Research Key Program [09JC1408000] and the Innovation Fund Project for Graduate Student of Shanghai (JWCXSL0801).

## References

1. Barabasi, A.L., Albert, R.: Emergence of scaling in random networks. *Science* 286(5439), 509–512 (1999)
2. Boccaletti, S., Latora, V., Moreno, Y., et al.: *Complex Networks: Structure and Dynamics*. *Physics Reports* 424(4,5), 175–308 (2006)
3. Guo, L., Xu, X.: *Complex networks*. Shanghai Scientific & Technological Education Publishing House, Shanghai (2006)
4. Arenas, A., Guilera, A.D., Kurths, J., et al.: Synchronization in complex networks. *Physics Reports* 469, 93–153 (2008)
5. Chen, G., Wang, X., Li, X., et al.: Some Recent Advances in Complex Networks Synchronization. In: Kyamakya, K. (ed.) *Recent Advances in Nonlinear Dynamics and Synchronization*, pp. 3–16. Springer, Heidelberg (2009)
6. Pecora, L.M., Carroll, T.L.: Master stability functions for synchronization coupled systems. *Physical Review Letters* 80(10), 2109–2112 (1998)
7. Wang, X., Chen, G.: Pinning control of scale-free dynamical networks. *Physica A* 310, 521–531 (2002)
8. Duan, Z., Wang, J., Chen, G., et al.: Stability analysis and decentralized control of a class of complex dynamical networks. *Automatica* (44), 1028–1035 (2008)
9. Jamshidi, M., Chen, Z., Huang, C. (translated): *Large-scale systems modeling and control*. Science Press, Beijing (1986)
10. Wang, Y.: *Large-scale systems—methodology and technology*. Tianjin University Press, Tianjin (1993)
11. Siljak, D.D., Zecevic, A.I.: Control of large-scale systems: Beyond decentralized feedback. *Annual Reviews in Control* 29, 169–179 (2005)
12. Bakule, L.: Decentralized control: An overview. *Annual Reviews in Control* 32, 87–98 (2008)
13. Zheng, W., Zhonghai, L., Siying, Z.: LMI approach of decentralized control and output tracking control for similar composite systems with mismatched interconnections. *Control and Decision* 15(4), 419–422 (2000)
14. Mao, C.-j., Yang, J.-h.: Decentralized Output Tracking for Linear Uncertain Interconnected Systems. *Automatica* 31(1), 151–154 (1995)
15. Boyd, S., Ghaoui, E., Feron, E., et al.: *Linear Matrix Inequalities in System and Control Theory*. SIAM, Philadelphia (1994)
16. Yu, L.: *Robust control—linear matrix inequality method*. Tsinghua University Press, Beijing (2001)
17. Huang, H., Han, J.: *Mathematical programming*. Tsinghua University Press, Beijing (2006)

# Manipulation of a Mobile Modular Manipulator Interacting with the Environment with the Assistance of Tactile Sensing Feedback

Jingguo Wang and Yangmin Li\*

Department of Electromechanical Engineering, University of Macau,  
Macao SAR, China  
[YMLi@umac.mo](mailto:YMLi@umac.mo)

**Abstract.** This paper presents a modeling method for a mobile modular manipulator (MMM) interacting with the environment, which is composed of a mobile base and a 5-DOF manipulator. A tactile sensor, which is installed at the tip of the end-effector of the mobile manipulator, is to detect the feedback information from the environment. The dynamic equations are formulated and the controller is proposed with both the redundancy resolution and manipulability measures. The tactile sensing-assisted strategies are proposed and combined into the control algorithm. A real mobile manipulator is built up to contact with the environment in our lab and the results show the effectiveness of the proposed method.

## 1 Introduction

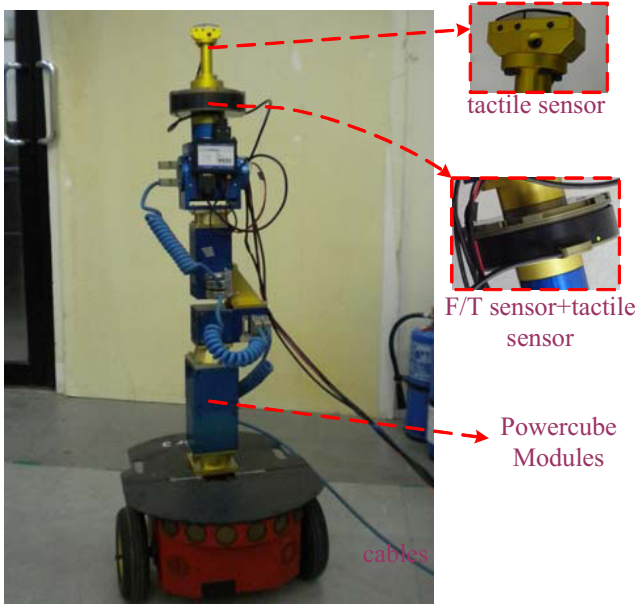
Considering that a mobile manipulator is able to perform manipulation tasks in a much larger workspace than a fixed-base manipulator, while a modular manipulator can change its configuration arbitrarily, we have designed a mobile modular manipulator (MMM) with the structure of a 5-DOF modular manipulator mounted on a mobile platform in this paper, and the platform is supported by two driving wheels and one caster wheel. Based on our previous works [1] [2], in addition to a six-axis force/torque sensor installed between the last joint and the end-effector, a new tactile sensor-based end-effector is fabricated and mounted on our MMM in order to provide tactile sensing feedback.

Like human sense of touch, tactile sensation plays an important role in robotic manipulation with high dexterity such as automatic grasping, edge tracking, rolling manipulation and so on [3]- [8]. For some complicated robotic manipulation in uncertain and dynamic environments, open loop control will not suffice, therefore it is essential to know whether contact has occurred and measure the parameters of contact between the sensor and an object like the spatial distribution of forces. The feedback of tactile sensing can be used to control force and torque at a specific contact location, which is necessary for manipulating objects and to control slippage [3].

The other parts of the paper are organized as follows. In Section 2, some related research works are studied, and the system description is given out in Section 3. In Section 4, the dynamic control of mobile modular manipulator is developed and the tactile

---

\* Corresponding author.



**Fig. 1.** The mobile modular robot

sensing strategies are considered and combined into the control algorithm. Section 5 is the part of the experiment. Some discussions and the conclusion remarks are obtained in the last section.

## 2 Related Research

How to properly and effectively coordinate the motions of the mobile platform and the upper manipulators becomes a critical problem. Yamamoto and Yun [9] [10] studied the effect of the dynamic interaction between the manipulator and the mobile platform of a mobile manipulator on the task performance, then they developed a nonlinear feedback that completely compensated the dynamic interaction. However, their dynamic model was somewhat inaccurate since the dynamic influence of the caster wheel was neglected. Yu and Chen [11] presented a general method to establish the dynamic equation of a mobile manipulator system by utilizing forward recursive formulation for open-loop multi-body systems. Tanner and Kyriakopoulos [12] studied the interaction forces between the manipulator base and the vehicle, which were assumed as virtual relative movements since these forces were useful in determining and maintaining contact stability of the wheels, although they did not contribute to the dynamics. Carriker et al. [13] decomposed the cost of point-to-point motion in three-dimensional Cartesian space into two components representing the qualitative difference between motion caused by the mobile base and motion caused by the manipulator system.

Cheng and Tsai [14] designed a tracking controller with fully dynamic compensation ability for the mobile manipulator system via Lyapunov stability theory. Kazuhiro

et al. [15] applied the impedance control method to control MR Helper and the control algorithm was implemented in the experimental system for handling an object in cooperation with the robot and the human. Bayle et al. [16] proposed a systematic modelling of the nonholonomic MMM to generalize the standard definition of manipulability to the case of mobile manipulators. Li and Liu [17] investigated the dynamic modeling and various control methods on the mobile manipulator, but redundancy is not considered. The redundancy resolution will be discussed in this paper.

### 3 System Description

The main configuration of the MMM is the five-DOF modular manipulator mounted on the mobile platform and the platform is supported by two driving wheels and one caster wheel as shown in Fig. 1 and Fig. 2. The mobile platform is the modified Pioneer mobile base (from MobileRobots) and the 5-DOF modular manipulator is made of three kinds of PowerCube modules (PR-070, PR-090 and PG-070 from Amtec robotics GmbH). The force sensor (FTCL-50-80 from SCHUNK GmbH & Co. KG) is mounted between the end-effector and the last joint, while the tactile sensor (DSA 9210 with controller DSA CON32-H from Weiss Robotics) having arc-like touching surface is installed at the tip of end-effector, providing a spatial resolution of 3.4mm with 70 sensor cells.

The main purpose of tactile sensor applied is to know whether contact has occurred, to determine the contact area and geometry from the sensor data and then to get the pressures of tactile cells. The working principle of the tactile sensors depends on an interface effect between metal electrodes and a conductive polymer covering the sensing electrodes, and the resistance between a common electrode and a sensor cell electrode is a function of the applied load and time [5]. The output voltage of a resistive sensor cell in the electrical circuit can be detected in real-time, and the pressure or force exerted on it will be obtained from the transducer characteristics.

### 4 Dynamic Control of the MMM with Tactile Sensing Feedback

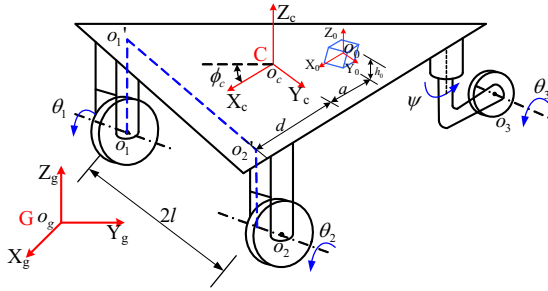
In the modeling of the MMM system, three assumptions should be adopted .

- The five-DOF modular manipulator is installed vertically on the planar platform.
- There are no slipping and no sideways between the wheels and the floor.
- Every link of the manipulator is rigid.

#### 4.1 Nonholonomic Constraints of Mobile Platform

The vehicle body moves horizontally in a nonholonomic fashion, subject to the combined effect of the forces exerted by the wheels and the forces exerted by the attached manipulator. The wheel forces are included in the form of reactions to the steering and traction torques exerted on the wheels [12].

The coordinate systems are defined as follows:  $G - X_g Y_g Z_g$  forms an inertial base frame(global frame)fixed on the motion plane and  $C - X_c Y_c Z_c$  is a frame fixed on the mobile platform.  $O_c$  is selected at the midpoint of the line segment connecting the two



**Fig. 2.** The configuration of mobile platform

fixed wheel centers; and  $Y_c$  is along the coaxial-line of the two fixed wheels. The heading angle  $\phi_c$  determines the posture of the mobile platform and  $\theta_1$  and  $\theta_2$  are the rotating angles of the left and right wheels as shown in Fig. 2

The mobile platform can either move in the direction of  $X_c$  or rotate about the  $Z_c$ , but the instantaneous motion in the direction  $Y_c$  is constrained because of the nonholonomic behavior of the wheels. Then the first constraint restricts the velocity of the mobile platform to be zero in the direction  $(\vec{o_1o_2})$  through two centers of driving wheels.

$$-\dot{x}_c \sin\phi_c + \dot{y}_c \cos\phi_c - d\dot{\phi}_c = 0 \tag{1}$$

The other two constraints are considered with the no-slipping of each rolling wheel in the forward directions, and the velocity of the driving wheels can be expressed as:

$$\dot{x}_c \cos\phi_c + \dot{y}_c \sin\phi_c - l\dot{\phi}_c = r\dot{\theta}_1 \tag{2}$$

$$\dot{x}_c \cos\phi_c + \dot{y}_c \sin\phi_c + l\dot{\phi}_c = r\dot{\theta}_2 \tag{3}$$

Let  $q_m = [x_c \ y_c \ \phi_c \ \theta_1 \ \theta_2]^T$  represent the basic coordinates ( $q_m$ ) of the MMM platform, so the above three constraints of velocities can be given in the matrix form:

$$A(q_m)\dot{q}_m = 0 \tag{4}$$

The independent velocities of two driving wheels are chosen as the generalized coordinates ( $\dot{q}_w$ ) because the nonholonomic constrained mobile platform owns only two-DOF under planar motion assumption and there is the following relation

$$\dot{q}_m = S(q_m)\dot{q}_w \tag{5}$$

where  $\dot{q}_w = [\dot{\theta}_1 \ \dot{\theta}_2]^T$ , and  $S(q_m)$  is a  $5 \times 2$  full rank matrix, whose columns are in the null space of  $A(q_m)$ .

With the additional DoFs of the upper modular manipulator, the extended coordinates of the whole system is composed of the wheeled mobile platform and the modular manipulator, whose form is  $q_e = [q_m^T \ q_1 \ \dots \ q_5]^T$ .

Now it can be parameterized in terms of the independent velocities  $q = [q_w^T \ q_1 \ \dots \ q_5]^T$

$$\dot{q}_e = S\dot{q} \tag{6}$$

where

$$S = \begin{bmatrix} S(q_m) & 0_{5 \times 3} \\ 0_{5 \times 2} & I_{5 \times 5} \end{bmatrix}$$

Combining the Eq. (4) and Eq. (5) together, the constraints can be easily obtained as:

$$A(q_m)S(q_m) = 0 \quad (7)$$

## 4.2 Dynamics Equation

Let  $q_e = [q_m^T \ q_1 \ \cdots \ q_5]^T$  denote the extended coordinates, with considering the external force/torques for the whole system with constraints ( $A(q_m)\dot{q}_m = 0$ ) in three dimensional operational space, the generalized dynamic equation of motion can be written as

$$\begin{aligned} M(q_e)\ddot{q}_e + C(q_e, \dot{q}_e)\dot{q}_e + G(q_e) &= E\tau + E_2F - A^T\lambda \\ A(q_m)\dot{q}_m &= 0 \end{aligned} \quad (8)$$

where  $M(q_e)$  is the nonsingular symmetric inertia matrix,  $C(q_e, \dot{q}_e)$  is the vector that implicitly includes centrifugal, Coriolis and viscous friction,  $G(q_e)$  is the gravity terms,  $E$  is the matrix maps the actuator torque to the joint torques,  $E_2$  is the matrix that maps the task space  $F$  to the joint space,  $A^T$  is the above constraint matrix and  $\lambda$  is the vector of Lagrange multipliers corresponding to the aforementioned constraint equations.

$$S^T(M(q_e)\ddot{q}_e + C(q_e, \dot{q}_e)\dot{q}_e + G(q_e)) = S^TE\tau + S^TE_2F - S^TA^T\lambda \quad (9)$$

With the equation of (7) and the derivative of equation of (6),

$$\ddot{q}_e = \dot{S}\dot{q} + S\ddot{q} \quad (10)$$

The motion equation of MMM can be written as

$$\bar{M}\ddot{q} + \bar{C}\dot{q} + \bar{G} = \bar{\tau} + \bar{F} \quad (11)$$

where  $\bar{M} = S^TMS$ ,  $\bar{V} = S^TVS + S^TM\dot{S}$ ,  $\bar{G} = S^TG$ ,  $\bar{\tau} = S^TE\tau$ , and  $\bar{F} = S^TE_2F$ .

## 4.3 Controller Design

A hybrid impedance controller in the task space is applied here, more details about redundancy resolution can be found in our related work [11],

$$u = \ddot{x}^d + B_d\dot{e} + K_d e + K_f(F_E^d - F_E) \quad (12)$$

where  $\ddot{x}^d$  is the desired acceleration,  $e$  and  $\dot{e}$  are the position error and velocity error,  $B_d$ ,  $K_d$  and  $K_f$  are the gains chosen appropriately with some techniques,  $F_E^d$  and  $F_E$  are the desired and actual forces of the end-effector. Then the desired impedance relation can be written as

$$\ddot{e} + B_d\dot{e} + K_d e = -K_f(e_f) \quad (13)$$

The aim of the end effector control law is to track the desired generalized task coordinates  $x_d$ , which includes position and force. With the decoupling of the task space and null space, the input torque should contain two parts, the torque of task space  $\tau_P$  and the torque of null space  $\tau_N$ ,  $\tau = \tau_P + \tau_N + C\dot{q} + G$ . Let the control law [21] be given as

$$\tau = J^T \{ \Lambda(u - \dot{J}\dot{q}) + F_e \} + C\dot{q} + G + \tau_N \quad (14)$$

where  $\Lambda \triangleq [JM^{-1}J^T]^{-1}$  denotes the pseudo kinetic energy matrix [20],  $C$  represents the coriolis and centripetal matrix and  $G$  denotes the gravity force vector.

This nonlinear control method separates the equations of motion into the task-space motion and the null-space motion and allows them to be controlled independently. Combining (12), (13) and (14), the task space joint torque has the form

$$\tau_P = J^T \{ \Lambda(\ddot{x}_d - \dot{J}\dot{q}) + B_d\dot{e}_p + K_d e_p + K_f e_f + F_e \} \quad (15)$$

A proper choice of  $\tau_N$  enables the manipulator to achieve secondary sub-task without affecting the task space dynamics, like force and trajectory tracking of the manipulator's end-effector. The null space torque vector will be obtained as

$$\tau_N = J_N^T \{ \Lambda_N(\ddot{x}_{Nd} - \dot{J}_N\dot{q}) + B_N\dot{e}_N \} \quad (16)$$

where  $\dot{e}_N$  is the velocity tracking error of the null space motion, which can be written as  $\dot{e}_N = \dot{x}_{Nd} - \dot{x}_N$ . One has the freedom to choose any control routine in the null-space [19]. The control scheme is designed as shown in Fig. 3.

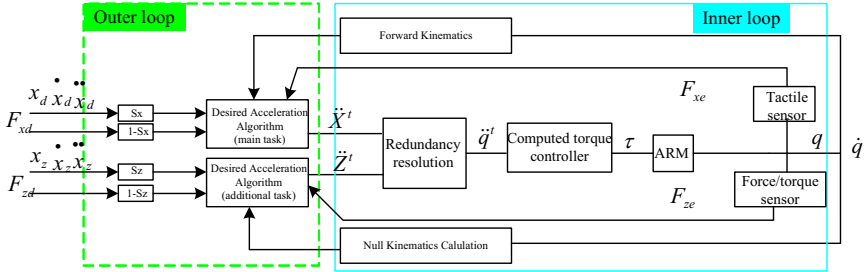


Fig. 3. The hybrid impedance control scheme

#### 4.4 Manipulability

The manipulability measure can be regarded as a distance measure of the manipulator configuration from singular ones at which the manipulability measure becomes zero [10]. At or near a singular configuration, the end point of the manipulator may not easily move in certain directions. The effort of maximizing the manipulability measure leads to keeping the manipulator configuration away from singularity, which is very important especially when a mobile manipulator is required to respond to motions whose range is unknown a priori. The manipulability measure can be defined by the Jacobian matrix of the manipulator as

$$\omega = \sqrt{\det(J(\theta)J^T(\theta))} \quad (17)$$

Since the task motion will be divided into the part of motion implemented by the mobile platform and the other part completed by the upper manipulator, the work [13] considers the problem of finding an optimal sequence of base positions and manipulator configurations for performing a sequence of tasks at specified locations in the workspace. The position of the manipulator end effector can be described by:

$$X_{tip} = X_{base} + X_{man}(q) \quad (18)$$

where  $X_{tip}$  and  $X_{base}$  are the locations of the end effector and the mobile base with respect to the global frame, and  $X_{man}(q)$  is a vector function that maps the manipulator joint variables  $q$  into the end-effector position relative to a Cartesian reference frame fixed on the mobile base.

To choose the most desirable positions and configurations for a sequence of  $n$  operations or tasks, a cost function is introduced in [13] which in some sense measures the energy or time required to transfer the system through the complete sequence of tasks. In general, the cost function is expressed as:

$$C_{tip} = \alpha C_{base} + \beta C_{man} \quad (19)$$

where  $C_{base} = \sum_{i=1}^n \delta_b(X_{base,i-1}, X_{base,i})$ ,  $C_{man} = \sum_{i=1}^n \delta_m(q_{i-1}, q_i)$ ,  $\delta_b$  and  $\delta_m$  are generalized distance functions representing the relative cost of moving the base and manipulator between two points respectively, and the parameters of  $\alpha$  and  $\beta$  have the relationship  $\alpha + \beta = 1$ .

With similar idea, the whole mobile manipulator manipulability is developed that considers the ability of generating velocities at the end-effector by acting on the sole arm or by acting on the whole system [16]. It is a convex combination of the arm manipulability measure and the mobile manipulator, a function has the form as:

$$P = \alpha \tilde{P}_{b+m} + (1 - \alpha) \tilde{P}_m \quad (20)$$

where  $\tilde{P}_m$  and  $\tilde{P}_{b+m}$  are the manipulability measures of the manipulator and the whole mobile manipulator respectively. The parameter  $\alpha$  has the value from 0 to 1. By this function, it is allowed to adapt the criterion to the mobile manipulator configuration or to the end-effector location.

#### 4.5 The Assistance of Tactile Sensing Feedback

In addition to the force-sensing feedback, tactile sensing can provide some useful and crucial information to the problems that the end-effector of the mobile manipulator is made a contact with the unknown environment.

Since tactile sensors determine various aspects of the environment through direct contact with objects within that environment [3], such sensors can be used to ascertain the distribution of forces over an area, and can also be used to sense shape parameters, location, and orientation of objects touching the sensor. Tactile sensors produce output as a digital image that always can be input to a conventional image processing system, which can be used for analyzing images from a vision camera. But in this paper, no



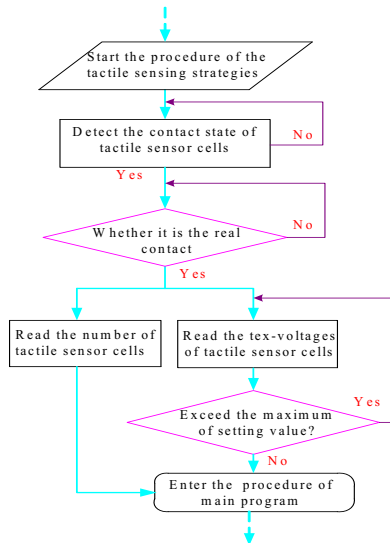
**Table 1.** Tactile Sensing Feedback Strategies

Strategy I:	To detect whether the contact occurred
Strategy II:	To judge the general contact information by the distribution of contact points
Strategy III:	To know the contact area from the number of tactile sensing cells
Strategy IV:	To avoid large pressures of some critical points

vision system is applied and the contact data (texel-voltage) of tactile sensor will be directly transferred into the control algorithm and some of the strategies are proposed in the table 1.

Safety is always the first to be considered if there is contact between the robot and human body, like massaging. The first strategy will ensure the safety and be in case of emergency just based on working state of sensor cells. From the second strategy, the contact type will be acquired from the pressure-value (or texel-voltage) distribution of sensor cells and contacted points can be known easily. In third strategy, the size of contact area can be calculated directly by summing up the number of contacted sensor cells, but note that the error will be decided by the size resolution of tactile sensor cell. Because the texel-voltage value of sensor cell can be obtained easily, some corresponding strategies should be made to avoid large pressures on some points and make quick response when they occur.

Therefore, based on the above tactile sensing strategies, the tactile sensing feedback will play an important role to contact control between the mobile manipulator and the environment. The flowchart of the tactile sensing procedure is shown in Fig. 4.



**Fig. 4.** The flowchart of tactile sensing strategies

## 5 Case Study

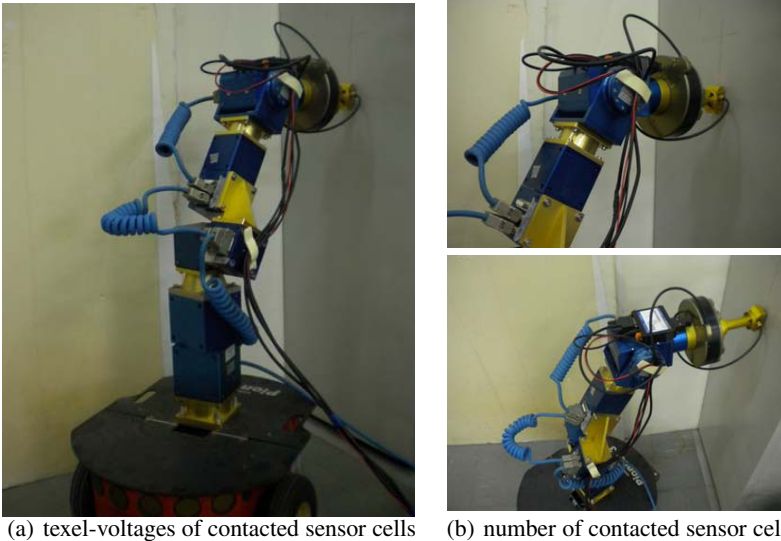
In order to demonstrate the effect of the tactile sensing feedback in the contact, experiments are made to test its force-tracking abilities on the planar surface of hard board.

### 5.1 System Parameters

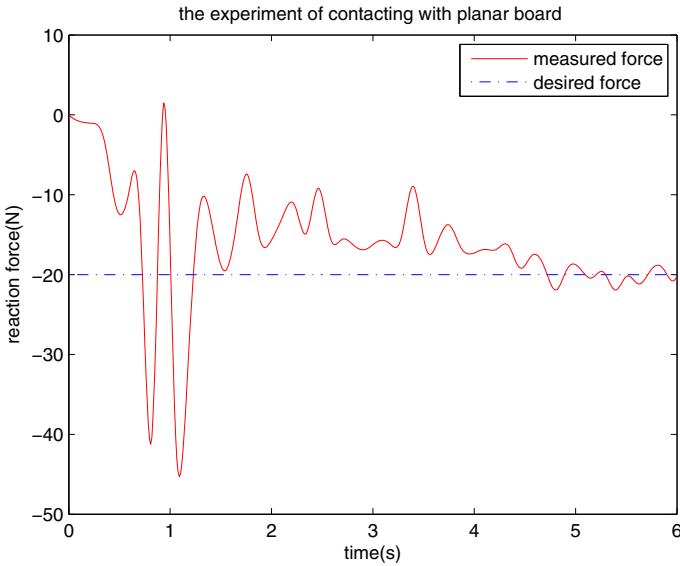
The mobile platform is the two-wheel driven Pioneer mobile base in our lab and its parameters can be found in [11]. A five-DOF manipulator actuated by PowerCube modules is equipped with force-torque sensor and tactile sensor. The link-lengths of the manipulator are  $a_1 = 0.375m$ ,  $a_2 = 0.337m$ ,  $a_3 = 0.308m$  and  $a_4 = 0.2851m$ , the masses of five modules are  $m_1 = 3.8kg$ ,  $m_2 = 6.6kg$ ,  $m_3 = 2.7kg$ ,  $m_4 = 1.8kg$  and  $m_5 = 1.05kg$  and the initial velocities will be set zero. The motor modules are communicated and controlled through CAN bus, while force and tactile sensing data is transmitted through RS-232 cables and then will be sent back to PC control. The weighting diagonal elements in the weighting matrices of the redundancy resolution have the values of  $w_e = 10$ ;  $w_n = 100$ ;  $w_v = 20$ . Other parameters' values for simulation are:  $M_d = 5$ ;  $B_d = 200$ ;  $K_d = 3000$ .

### 5.2 Following the Trajectory on the Surface of Planar Board

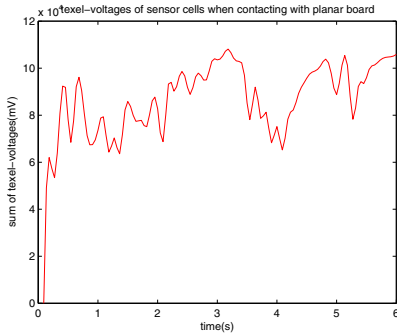
The mobile manipulator is planned to follow the planar trajectory on a planar board as shown in Fig. 5. The whole manipulation is composed by the motion of mobile base and the motion of the upper manipulator as discussed above. A target force 20N is



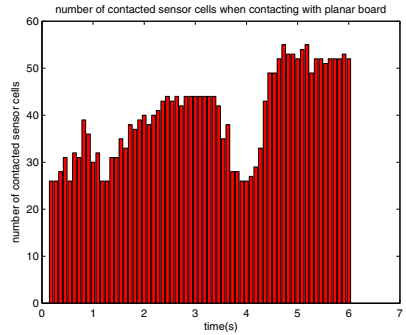
**Fig. 5.** Two experiments of contacting with foam plate and hard board respectively



**Fig. 6.** The reaction force measured in the simulation



(a) texel-voltages of contacted sensor cells



(b) number of contacted sensor cells

**Fig. 7.** Experiment of contacting with planar hard board

desired to be followed, which is orthogonal to the plane of the trajectory. The actual force measured through the force sensor is recorded as shown in Fig. 6. At the same time, the texel-voltages of all the tactile sensor cells are recorded as shown in Fig. 7(a). Since the contacted area can be calculated through small cells, the number of contacted sensor cells are also recorded as shown in Fig. 7(b). From Fig. 7(a) and Fig. 7(b), it is not difficult to find that the whole texel-voltages of tactile sensor cells have almost the similar change as the number of contacted cells. It can be known that most of the contacted cells don't work in the state of saturation.

## 6 Discussion and Conclusion

### 6.1 Discussion

In addition to the force information read from the force/torque sensor, the feedback of tactile sensing still plays an important role in the dynamic control, although it is not the direct element in the dynamic equation. The tactile sensing can provide many vital strategies in the whole control algorithm. However, the proposed method of tactile sensing in the paper has limited use in the case of big contacted force, which means that the contacted tactile sensor cells almost come into saturation state so they can not provide the valuable information. The exact reason is that the texel-voltage of sensor cell will not change in case of the saturation state.

### 6.2 Conclusion

This paper proposed a mobile modular manipulator with the tactile sensor installed at the tip of the end-effector, which is built up by a mobile base and a five-DOF manipulator. The dynamic equations are formulated based on the Lagrangian equations with the nonholonomic constraints of the mobile platform. The manipulability is considered and the tactile sensing strategies are proposed and combined into the control algorithm. A real experiment is made to let the end-effector to contact with the planar board, some valuable results are obtained.

## Acknowledgement

This work is supported by Macao Science and Technology Development Fund under Grant no. 016/2008/A1.

## References

1. Wang, J.-G., Li, Y.M.: Dynamic Control and Analysis of a Nonholonomic Mobile Modular Robot. In: IEEE Int. Conf. on Intelligent Robotics and Application, Singapore, pp. 776–791 (2009)
2. Wang, J.-G., Li, Y.M.: Dynamic Modeling of a Mobile Humanoid Robot. In: IEEE Int. Conf. on Robotics and Biomimetics, Bangkok, Thailand, pp. 639–644 (2009)
3. Tegin, J., Wikander, J.: Tactile sensing in intelligent robotic manipulation: A review. *Industrial Robot: An Int. J.* 32(1), 64–70 (2005)
4. Zhang, H., Chen, N.: Control of contact via tactile sensing. *IEEE Trans. on Robotics and Automation* 16(5), 482–495 (2000)
5. Schmid, A.J., Gorges, N., Göger, D., Wörn, H.: Opening a door with a humanoid robot using multi-sensory tactile feedback. In: IEEE Int. Conf. on Intelligent Robots and Systems, Pasadena, CA, USA, pp. 285–291 (2008)
6. Lo, W.T., Shen, Y.T., Liu, Y.H.: An integrated tactile feedback system for multifingered robot hands. In: IEEE Int. Conf. on Intelligent Robots and Systems, Hawaii, USA, pp. 680–685 (2001)
7. Iwata, H., Sugano, S.: HumanCrobot-contact-state identification based on tactile recognition. *IEEE Trans. on Industrial Electronics* 52(6), 1468–1477 (2005)

8. Mukai, T., Onishi, M., Odashima, T., Hirano, S., Luo, Z.W.: Development of the tactile sensor system of a human-interactive robot RI-MAN. *IEEE Trans. on Robotics* 24(2), 505–512 (2008)
9. Yamamoto, Y., Yun, X.: Effect of the Dynamic Interaction on Coordinated Control of Mobile Manipulators. *IEEE Trans. on Robotics and Automation* 12(5), 816–824 (1996)
10. Yamamoto, Y., Yun, X.: Coordinating Locomotion and Manipulation of a Mobile Manipulator. *IEEE Trans. on Automatic Control* 39(6), 1326–1332 (1994)
11. Yu, Q., Chen, I.-M.: A General Approach to the Dynamics of Nonholonomic Mobile Manipulator Systems. *J. Dyna. Sys. Meas. and Contr.*, Transactions of the ASME 124, 512–521 (2002)
12. Tanner, H.G., Kyriakopoulos, K.J.: Mobile manipulator modeling with Kane's approach. *Robotica* 19, 675–690 (2001)
13. Carriker, W.F., Khosla, P.K., Krogh, B.H.: An Approach for Coordinating Mobility and Manipulation. In: *IEEE Int. Conf. on Systems Engineering*, pp. 59–63 (1989)
14. Cheng, M., Tsai, C.: Dynamic Modeling and Tracking Control of a Nonholonomic Wheeled Mobile Manipulator with Two Robotic Arms. In: *IEEE Conf. on Decision and Control, Hawaii*, vol. 3, pp. 2932–2937 (2003)
15. Kazuhiro, K., Manabu, S., Norihide, K.: Mobile Robot Helper. In: *IEEE Int. Conf. on Robotics and Automation, San Francisco*, pp. 583–588 (April 2000)
16. Bayle, B., Fourquet, J.Y., Renaud, M.: Manipulability of Wheeled Mobile Manipulators: Application to Motion Generation. *Int. J. Robotics Research* 22, 565–581 (2003)
17. Li, Y.M., Liu, Y.G.: Dynamic Modeling and Adaptive Neural-fuzzy Control for Nonholonomic Mobile Manipulators Moving on a Slope. *Int. J. Contr. Autom. Systems* 4(2), 1–7 (2006)
18. White, G.D., Bhatt, R.M., Krovi, V.N.: Dynamic Redundancy Resolution in a Nonholonomic Wheeled Mobile Manipulator. *Robotica* 25, 147–156 (2007)
19. Wang, J.-G., Li, Y.M.: Impedance control of a spatial redundant manipulator used for relaxing muscle fatigue. In: *IEEE Int. Conf. on Mechatronics and Automation, Changchun, Jilin, China*, pp. 2799–2804 (2009)
20. Khatib, O.: A Unified Approach for Motion and Force Control of Robot Manipulators: The Operational Space Formulation. *IEEE J. of Rob. and Auto.* 3, 43–53 (1987)
21. Nemeč, B., Zlajpah, L.: Null space velocity control with dynamically consistent pseudo-inverse. *Robotica* 18(5), 513–518 (2000)

# An RTC-Based Intelligent Programming System for Service Robots

Masaru Adachi<sup>1</sup>, Zhen Zhang<sup>2</sup>, Peihua Chen<sup>2</sup>, and Qixin Cao<sup>2</sup>

<sup>1</sup> Yaskawa Electric Corporation, 12-1, Ohtemachi, Kokurakita-ku, Kitakyushu, Fukuoka 803-8530, Japan

<sup>2</sup> Research Institute of Robotics, Shanghai Jiao Tong University, Shanghai, 200240, China

**Abstract.** A service robot is a complicated system integrating multi-disciplinary technologies, and robotic graphical programming is one of the key technologies for intelligent service robots. Based on the modularization and standardization of robot technology, this paper proposes an RTC (Robot Technology Component) based distributed structure framework robot system. In addition, this paper develops RTC functional components and establishes a 3D simulation and graphical programming system for intelligent service robots. A 3D simulation validation experiment is also presented, demonstrating that SmartPal dual-arm service robots execute pick & place tasks. The experiment shows that this intelligent programming system is an excellent robotic development tool. It has strong advantages in both its expansibility and integration ability, and it provides a convenient way for reducing the development costs of intelligent service robots.

**Keywords:** service robot; 3D simulation; graphical programming; RTC (robot technology component).

## 1 Introduction

With the continuous development of robotics technology, robot systems based on functional components have gradually become the focus of robot system research and industrial development. Research on robot middleware technology helps change the perspective of design and development of robot systems from general modeling to modular modeling, with the reuse of hardware and software components. Robot middleware technology helps to lower the threshold of robot design, simplify the design process, and reduce the costs of robot system research and development, and is attracting more and more academic attention.

There are several examples of this trend towards modularization, one of which is the OROCOS project. Started in Europe, it aims to develop a common modular architecture for robot control. The OROCOS project designs and develops the four C++ development libraries: a real-time tool set, a set of kinematics and dynamics algorithms, a Bayesian filtering database, and an OROCOS component library for building a common modular framework [1-3]. OPEN-R is the standard interface for the entertainment robot system that SONY is actively promoting. This interface greatly expands the capabilities

of entertainment robots like AIBO, SDR-3X and QRIO [4]. One further example is, the Industrial Robot Committee of the Japan Robot Association, Tokyo University and the Manufacturing Science and Technology Center's joint research project - OpenHRP (Open architecture Humanoid Robotics Platform), aims to develop a common modular platform for humanoid robot simulation and actual control, mainly for the simulation and control of kinematics planning, collision detection, dynamics simulation, etc., for various types of humanoid robots [5-7]. Finally, in advancing the features of robot system development and modular design, Japan introduced the robot technology component (RTC). It is based on the CORBA specification with some simplification and complementation, using RT (Robot Technology) middleware to integrate different RTCs to speed up the integration of modular components of a robot system [8-11].

Following this trend, this paper explains the development of a 3D simulation and graphical programming system based on RTC technology for intelligent service robots and takes both modular and standardized design ideas for indoor service robots into consideration. Through this system, we realized interactive robot navigation, and off-line programming and simulation of pick & place applications, thereby verifying the practicality and simplicity of RTC technology.

## 2 System Architecture

The system is composed of three parts (RT components, 3D tools and a graphical programming tool) as shown in Fig. 1. Various software modules to control the robot are constructed by the functional processing module as the RT component. The graphical programming tool describes the application by controlling these RT components. In addition, the tools that help achieve the off-line programming in a 3D virtual environment are 3D Map Editor and 3D Viewer. Sections 3, 4 and 5 explain the contents of each software module.

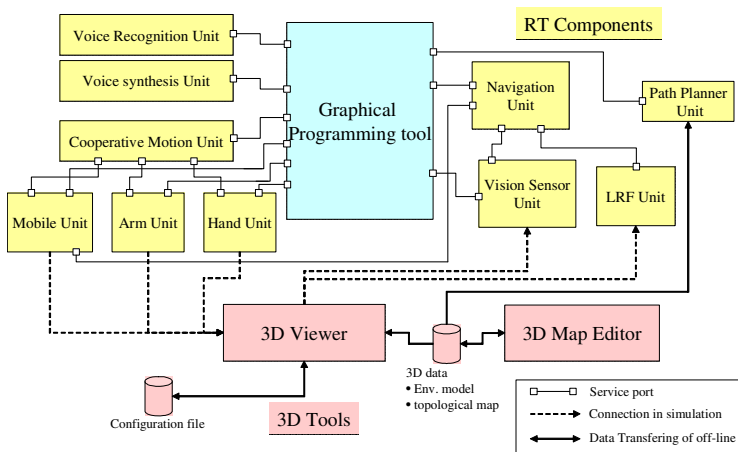


Fig. 1. System Block Diagram

Prior to the use of RT middleware, we modulated robot control software in the functional processing module. Also, we had used CORBA to call the function of the module remotely from the outside. These were called the RT unit. Similarly, the RT middleware [12] also offers a common platform where the construction of various robotic systems is enabled by freely combining a variety of robot elements through the communication network.

Because the OpenRTM-aist [11] (one of the RT middleware implementations) that we had used adopted CORBA in the communication part, it was highly compatible with the RT unit. Therefore, it was easy to modify the RT unit to the RT component. In this paper, the software module corresponding to the RT middleware will be called the RT Component or RTC (Robot Technology Component).

The features of this architecture are the following three points.

- Interconnection and reuse is easy because the robot control software is developed as RTC. Robot systems in various forms can be constructed by combining necessary RTCs from among existing RTCs.
- The readability and debugging operations are easy, because a graphical programming tool is used for the application development.
- It can develop applications and RTCs itself in a 3D virtual environment, even if a real robot system is not used, by using 3D Tools and a virtual RTC that emulates a real sensor and real motion. In addition, the application developed by the virtual environment can be executed as it is in a real robot system.

### 3 RT Components

#### 3.1 Arm Unit RTC and Mobile Unit RTC

An Arm Unit is a 7 DOF manipulator that Yaskawa developed for service robots. It is possible to operate it by giving commands only from upper layer RTC because it provides basic motion operation instruction known as PTP (Point to Point) and CP (Continuous Path). An Arm Unit RTC offers about 30 commands in the service port. Table 1 shows the representative commands.

**Table 1.** Representative commands of Arm Unit RTC

Command Name	Brief overview
moveJointAbs	In the joint space, the interpolation operation is done to the terminal position in which each axis angle was specified.
moveLinearAbs	In the cartesian space, the interpolation operation is done to the terminal position in which position and posture were specified.
pause	The operation of all axes is stopped temporarily.
resume	The temporary stop is released, and the operation of all axes is restarted.



A Mobile Unit is an omnidirectional service robot that Yaskawa developed. It is possible to operate it by giving commands only from the upper layer RTC because it provides basic motion operation instruction. A Mobile Unit RTC offers about 40 commands in the service port. Table 2 shows the representative commands.

**Table 2.** Representative commands of Mobile Unit RTC

Command Name	Brief overview
moveLinearAbs	It moves to the target position specified in the absolute coordinate system.
moveCruiseAbs	It moves while drawing a smooth curve trajectory from present position to the target position.
getState	The state of the mobile unit is returned.
getPosition	The current feedback position information is returned.

### 3.2 Vision Sensor RTC

In investigating the various parameters of a real vision sensor, it is proposed that the vision sensor consists of the following controllable parameters: the horizontal field of view (fov<sub>x</sub>), the window aspect ratio (aspect), the focal length and the sensor's position and orientation in the scene.

In addition to these parameters, the virtual vision sensor component has another two special parameters, which are before and after the shear plane (zNear, zFar). These two parameters are used to determine which objects in the range are rendered onto the canvas, and objects too far or too close from the vision sensor will be ignored. A Vision Sensor RTC offers 4 commands in the service port. Table 3 shows the representative commands.

**Table 3.** Representative commands of Vision Sensor RTC

Command Name	Brief overview
getPos	Get the vision sensor's posture
setPos	Set the vision sensor's posture
getImage	Get an image shot
getPosition	Get the position of the specified target

The Data port of the vision sensor component is mainly for data output, and sending image data to a particular external connection. First, the component collects images from the virtual environment. Considering the requirements of a real-time vision sensor, this paper adopts the method of acquiring the real-time image on a designated area of the screen's pixels. Using the CreateScreenCapture method of the Robot class, this component obtains the screen pixels of the virtual sensor's window area, and saves it in the image buffer. After the image acquisition, the vision sensor RTC changes the image into a two-dimensional array, using the first data for saving the image width and the following data for storing the pixels of the image and packages for the CORBA

interface function - `getImage ()`. This function becomes the data output port of the vision sensor RTC, and other components can access the image data through this port.

The virtual vision sensor component has the function of image processing and recognition. Through the service port, it provides specific service resources to the outside, for example by providing the identification and location of a cup using the feature of color. The external components need to provide the color characteristics of a target to the service port, and then the vision sensor RTC recognizes the target and calculates its position according to its features, and returns the results to the external components.

The vision sensor RTC has a configuration interface. The interface provides a dynamic external call, modifies the internal logic parameters, and provides a CORBA interface function - `setPose ()`. Through the interface, other components can dynamically adjust the 6 DOF of the virtual vision sensor in the virtual environment, can display the scene from a specific view, and can obtain the scene image from a specific perspective view.

### 3.3 Path Planner RTC

In order to achieve autonomous locomotion, robots must make path planning and motions based on the information of a global map. For indoor service robots, the global map is known, Therefore, we developed a path planner RTC. Using the global geometric map, the path planner RTC generates a topological map, and then executes path planning according to the specific task.

The main steps are as follows:

1. Change geometric maps into binary grid maps
2. Generate a reachable path for the robot based on the Voronoi algorithm
3. Create topological nodes according to the task and based on the reachable path
4. According to A\* algorithm and the Artificial Potential Field Method, the robot obtains the best path.

A Path Planner RTC offers about 5 commands in the service port. Table 4 shows the commands.

**Table 4.** Representative commands of Path planner Unit RTC

Command Name	Brief overview
<code>getPath</code>	Get the topological path data
<code>computePath</code>	Compute the topological path
<code>getNextPath</code>	Get next node based on the result by <code>computePath</code> command.

## 4 3D Tools

### 4.1 3D Map Editor

3D map editor is an interactive map creating and editing tool that can help complete the task of map-based autonomous locomotion. This tool creates, edits, and displays the 3D

virtual environment. The information about the work environment can be saved in an XML (eXtensible Markup Language) format for further reading and editing, according to a certain specification (such as DTD or Schema) [13].

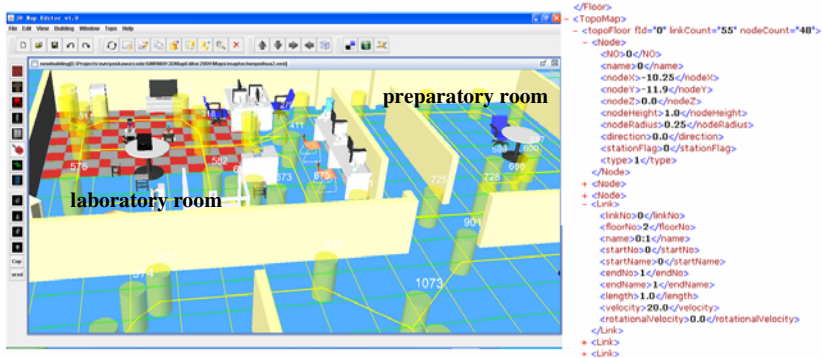


Fig. 2. Create and save of a topological map

In addition, the 3D map editor tool also has a built-in topological map, with the functions of automatic generation and manual edit, and can be saved in specified forms into the above work environment. The topological map can provide data and service to the robot control and path-planning components. In a work environment such as the one shown in Fig. 2 (left) for example, the corresponding topological map data is shown in Fig. 2 (right).

## 4.2 3D Viewer

The 3D robot model component adopts Java 3D technology, using the VRML (Virtual Reality Modeling Language) model [14] generated by UG, Pro/E, SolidWorks and other 3D modeling software to make the parameterized assembly in the 3D virtual environment in accordance with the robot's scene graph, and makes the parameters configurable, as shown in Fig. 3. After virtual assembly and design of the appropriate control panel, or some other corresponding interfaces, the 3D robot model component assesses the motion control of each robot, and makes a simulation to demonstrate various movements according to the users' requests.

Fig.3 shows the structure of the 3D robot model component, in which the parameterized robot assembly and its configuration structure are proposed and then realized as shown in Fig.4. At the same time, for different types of robots, it provides the uniform external CORBA interface function - setPos (), and other RTCs can control each robot through this RTC's service.

Once the service robot simulation is complete, the 3D viewer tool provides an integrated three-dimensional simulation platform that can read the map files provided by the 3D map editor tool, load in 3D robot models provided by the 3D robot model component, identify and process the vision data fed back by the vision sensor RTC, and

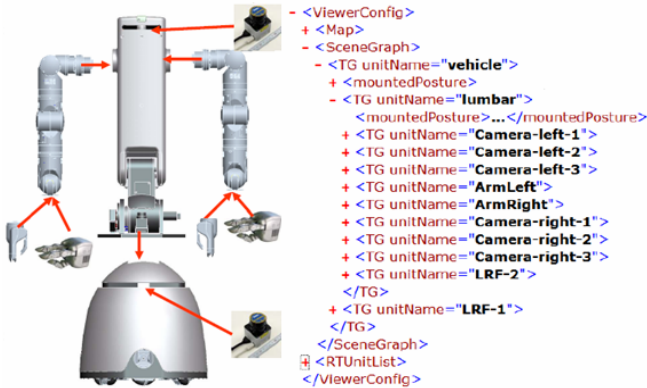


Fig. 3. Virtual assembly of 3D robot model

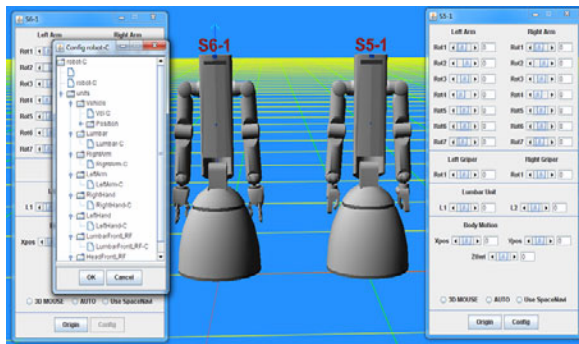


Fig. 4. An example for 3D robot modeling and configuration

carry out the robot control strategies and path planning tasks according to the robot control and path planner RTCs. Thus, the 3D viewer tool can conduct a systematic, real-time, accurate three-dimensional simulation.

## 5 Graphical Programming Tool

The graphical programming tool is used to specify the robot’s task by using a list of motion icons (Fig. 5). Users can edit each robot’s joint angles in the teaching box to define a motion, and then check the result in its 3D viewer. Once an icon is programmed, it can be saved into an icon list. Using this tool, the user can program the robot to complete motion tasks.

As a user-programming tool, it provides a user interface (UI) to control RTCs and integrates them based upon the user’s commands. The graphical programming tool controls sensor components (vision and laser component) so that they send their data periodically, runs the arm unit and mobile unit, and updates the status data in the 3D viewer tool.

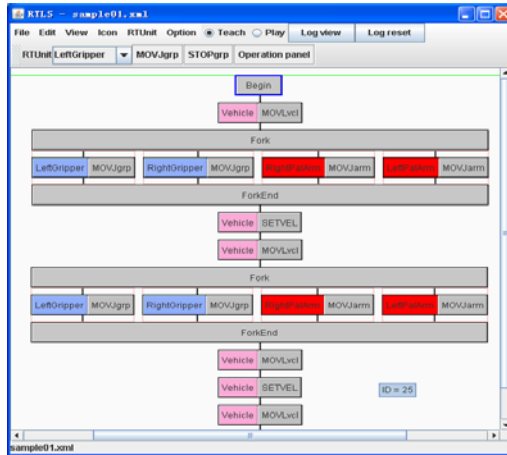


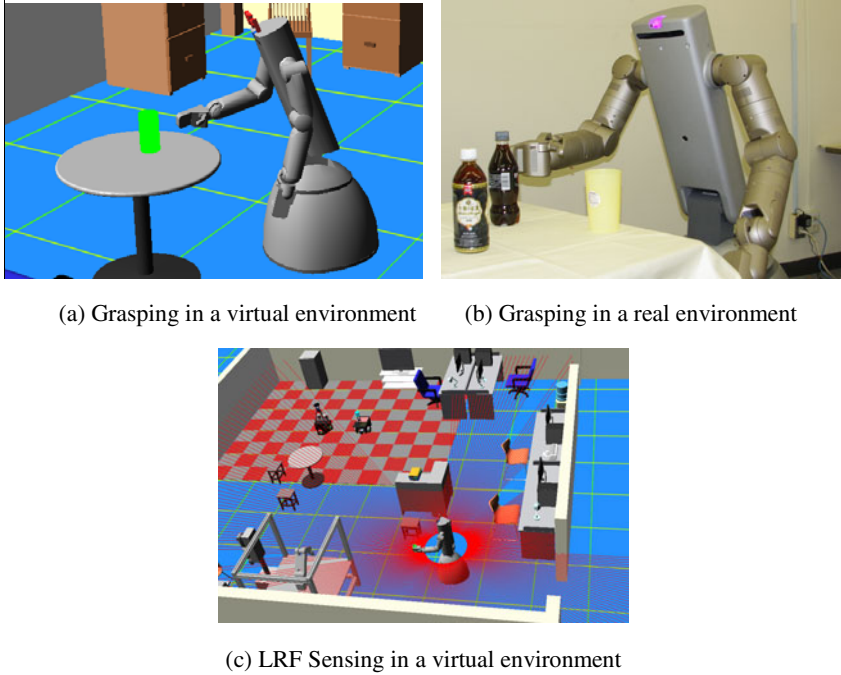
Fig. 5. Graphic programming tool

## 6 Experiment and Simulation

The application of pick & place was constructed by using the above-mentioned RTCs, 3D tools, and graphical-programming tool. The experiment environment is composed of two rooms (laboratory room and preparatory room) as shown in Fig. 2. Two or more desks and PCs are put in the laboratory room. In the preparatory room, one table is put near the center of the room. The robot uses SmartPal (dual-arm mobile type). The list of RTCs to control the robot are shown in Table 5. The corresponding block diagram is as shown in Fig. 1.

Table 5. A list of RTCs for experiment

RTC Name	Brief overview
Voice Recognition Unit	To talk with the person, voice recognition is executed.
Voice Synthesis Unit	To talk with the person, speech is executed.
Cooperative Motion Unit	The arm, the waist, and the movement truck are cooperatively made to work.
Mobile Unit	The mobile function of the robot is controlled.
Arm Unit	The manipulation function of the robot is controlled.
Navigation Unit	The migration path of the robot is controlled, and the obstacle avoidance is executed.
Path Planner Unit	The migration path is generated.
Vision Sensor Unit	Positional posture of the object is measured.
LRF Unit	The environment is measured by two dimensions.



**Fig. 6.** 3D simulation for intelligent service robots based on RTC

An application program was developed that achieved the following operation.

1. The robot moves from the laboratory room to preparatory room.
2. The robot grasps the glass on the table (Fig. 6 (a) and (b)).
3. The robot moves from the preparatory room to laboratory room.
4. The robot puts the glass on the table.

In Step 1 and Step 3, the route is derived by using the Path Planner Unit, and the Navigation Unit moves the robot to the target room based on the result. The Navigation Unit avoids the obstacle based on the data of the LRF Unit (Fig. 6 (c) ). In Step 2 and Step 4, the position of the glass is measured with the Vision Sensor Unit, and the hand and the arm are operated based on the result, and the glass is grasped/put.

The above-mentioned application was first constructed in a virtual environment, it was debugged next in a real environment and then completed. The benefit of using this process is that the application program could be initially created and tested all within a virtual environment, saving time and resources. Then, only the final test run would have to be carried out in the real environment. (A real robot-debugging step is needed in order to adjust for unexpected factors that the robot comes in contact with in a real environment.)

## 7 Conclusion

In this paper, a graphical programming and 3D simulation system based on RTC for service robots is presented. Through this experiment, it was confirmed that this system has the following advantages.

Service robots can be built in a short time by reusing many RTCs that were developed at Yaskawa and Shanghai Jiao Tong University. A Service robot is composed of many RTCs. The integration application can be efficiently developed by the graphical programming tool. At Shanghai Jiao Tong University, SmartPal is used in the 3D virtual environment of this system, and it is seen that RTCs and applications can be developed based on this. In Yaskawa, the application program completed by the virtual environment can control real SmartPal.

In summary, by examining the developing process and the experiment results of the intelligent service robots' graphical programming and 3D simulation system based on RTC, it is seen that RTC technology based on the modular design of robot architecture greatly reduces the development costs and the threshold of robot development, and has a large practical significance. Future plans include the upgrade of each RTC, and to develop functions such as dynamic simulation and collision detection checks with the 3DViewer.

\* SmartPal is a registered trademark of Yaskawa Electric Corporation.

**Acknowledgments.** This research was performed at the Research Institute of Robotics, Shanghai Jiao Tong University and Yaskawa Electric Corporation. We wish to express our gratitude to everybody who cooperated in the program development and experiment.

## References

1. Open System Architecture for Controls within Automation Systems EP 6379 & EP 9115 OSACA I & II Final Report (version 1.5) (April 30, 1996)
2. <http://www.osaca.org>
3. Bruyninckx, H.: Open robot control software: the OROCOS Project. In: Proc. Int. Conf. Robot. Autom., Seoul, Korea, pp. 2523–2528 (2001)
4. Sabe, K.: Open-R: An Open Architecture for Robot Entertainment. In: AIM 2003 Workshop: Middleware Technology for Open Robot Architecture (2003)
5. Fumio, K., Hirohisa, H., Shuuji, K.: OpenHRP: Open Architecture Humanoid Robotics Platform. I. J. Robotic Res. 23(2), 155–165 (2004)
6. Kanehiro, F., Fujiwara, K., Kajita, S.: Open Architecture Humanoid Robotics Platform. OpenHRP 21(7), 785–793 (2003)
7. Kanehiro, F., Fujiwara, K., Kajita, S.: Open Architecture Humanoid Robotics Platform. In: Proceedings of the 2002 IEEE International Conference on Robotics & Automation, Washington, DC (May 2002)
8. Ando, N., Suehiro, T., Kitagaki, K.: RT-middleware: distributed component middleware for RT (robot technology). In: Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2005), pp. 3933–3938 (2005)
9. Makoto, M.: Robot Technology (RT) Trend and Standardization. In: IEEE Workshop on Advanced Robotics and its Social Impacts (2005)

10. Hada, Y., Jia, S., Takase, K., et al.: Development of Home Robot Integration System Based on Robot Technology Middleware. In: The 36th International Symposium on Robotics (ISR 2005), TU4H6, Japan (2005)
11. RT-Middleware Project Home Page, <http://www.is.aist.go.jp/rt/> (cf. 2007.4.10)
12. Documents Associated With Robotic Technology Component (RTC), Version 1.0, <http://www.omg.org/spec/RTC/1.0/>
13. <http://www.w3school.com.cn/>
14. <http://www.w3.org/MarkUp/VRML/>



# Marine Engine State Monitoring System Using Distributed Precedence Queue Mechanism in CAN Networks

Hyun Lee<sup>1</sup>, Dong kyu Yi<sup>1</sup>, Jun seok Lee<sup>2</sup>, Gye-do Park<sup>2</sup>, and Jang Myung Lee<sup>2</sup>

<sup>1</sup> heriana, <sup>2</sup> Department of Electronics Engineering, Pusan National University  
San 30 Jangjeon-dong Kumjeong-ku, Busan, 609-735, Korea  
hlee@pusan.ac.kr, subtract@nate.com, sukidda@pusan.ac.kr,  
smelrose@pusan.ac.kr, jmlee@pusan.ac.kr

**Abstract.** This paper proposes a marine engine state monitoring system using distributed precedence queue mechanism which collects the state of bearings, temperature and pressure of engine through the CAN network. The CAN is developed by Bosch Corp. in the early 1980'. The data from various sensors attached in the marine engine are converted to digital by the analog to digital converter and formatted to fit the CAN protocol at the CAN module. All the CAN modules are connected to the SPU module for the efficient communication and processing. This design reduces the cost for wiring and improves the data transmission reliability by recognizing the sensor errors and data transmission errors. The distributed precedence queue mechanism is developed for the performance improvement of the marine engine system, which is demonstrated through the experiments.

**Keywords:** CAN (Controller Area Network), ESMS(Engine State Monitoring System), DPQ (Distributed Precedence Queue).

## 1 Introduction

Inspecting the partial elements of the engine of the Marine to check the status of the engine and warning is very important. But there exist various kinds of obstacles to check the all of these parameters. To disassemble and inspect the cylinders, bearings, valves, status of the gas in the vales, and coolant oil after assemble of the engine takes lots of time and money. Furthermore, inspecting the engine by the human-being is dangerous for the engineers. For these reasons, inspecting and monitoring the status of the engine to prevent the unpredicted accident is important system, so this system must contains diagnostic and protection circuit, which could be one of the best solutions for the shut-down of the system, to guarantee the reliability.

Structural techniques for minimizes external environmental effects and Electrical circuit Program design techniques protects from external factors which affects precise system operation are needed.

Various other kinds of techniques are needed, for example, ubiquitous sensor networking of the vessel, simplification the electrical wiring using wireless communication, remote control system, engine control system, and integration of the communication. The main principle of the bearing state monitoring system could be realized by inspecting the distance between huge two-stroke cross header of diesel engine and displacement sensor, and estimate the status of the abrasion on the bearing surface. CAN communication make real time monitoring possible for the temperature of the cylinder gas, cooling fresh water, main bearing, and PCO non-flow switch signals. To estimate the abrasion status and temperature of the several types of engine bearing, which are crosshead bearing, crank bearing, main bearing, can realize the real-time monitoring system for the engine and prevent the unexpected accident. This paper proposed a new engine state monitoring system related to the status of the engine control room.

## 2 Basic CAN Protocol

CAN is based on a CSMA/CD channel access technique, modified to enforce a deterministic resolution of collisions on a network which uses a priority scheme based on the identifiers of the exchanged objects. The CAN protocol adopts a layered architecture which is based on the OSI reference model, even though it is not fully OSI compliant. As with other networks conceived for the factory automation environment, it relies on a reduced protocol stack, consisting only of the following three communication layer:

- Physical layer
- Data link layer
- Application layer

The network, transport, session and presentation layers have eliminated so as to obtain a streamlined protocol which provides shorter response times. It is worth noting that up to now only the physical and the data link layers have been standardized, however a standard proposal was recently completed for the application layer.

In this paper interest is focused mainly on the data link layer. In fact most of the features of CAN which concern topics such as the sharing of the bandwidth among

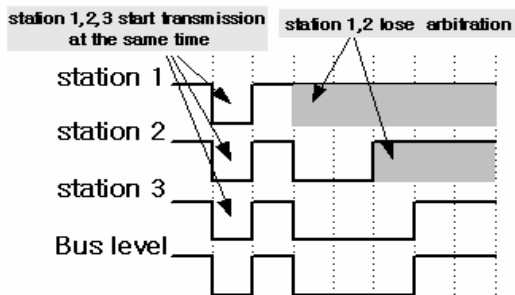


Fig. 1. CAN arbitration phase

the different stations and the access delays they experience depend on the mechanisms adopted at this level. According to a well-known model developed for LANs, the data link layer is, in turn, divided into two sub-layers, namely:

- Medium access control (MAC);
- Logical link control (LLC).

### 3 Distributed Precedence Queue (DPQ) Mechanism

The CAN implicitly assigns to each object exchanged in the network a priority that corresponds to the identifier of the object itself. Even though this mechanism enforces a deterministic arbitration that is able to resolve any conflict that occurs when several nodes start transmitting at the same time, it is clearly unfair. If many nodes are connected in the network, nodes that are of low priority rank can continuously lose a transmission opportunity. That is, if high priority objects transmit continuously, finally a low priority object can miss an important message which is relatively unimportant compared to that of a high priority object.

Accordingly, a mechanism that uses a relative priority according to the consideration of low priority nodes is necessary although the CAN implicitly assigns a priority. Fair behavior, which for example enforces a round-robin policy among different stations, has to be guaranteed to all the objects exchanged at a given priority level.

In this paper, it is shown that this kind of behavior can be obtained by slightly modifying the frame acceptance filtering function of the LLC sub-layer. In particular, only the significance of the identifier field in the transmitted frame has to be modified in some way. The resulting arbitration mechanism is able to enforce a round-robin policy among the stations that want to transmit a message on the bus, and provides two levels of priority for the frame transmission services. Little or nothing has to be changed at the MAC level; and in this way it is possible to reuse the same electronics components developed for the implementation of the standard CAN protocol.

#### 3.1 DPQ (Distributed Precedence Queue) Basic Principle

The basic idea of this CAN fairness control mechanism that is to insert into a global queue all of the nodes that want to transmit over the shared medium. For Node C, of

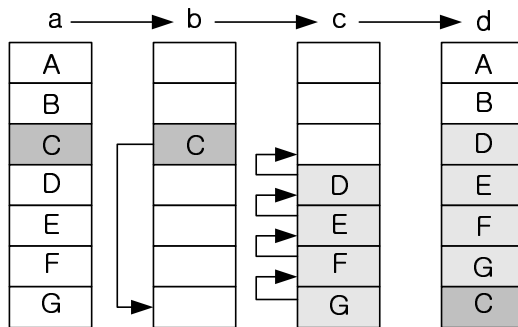


Fig. 2. Generation of a precedence queue in DPQ mechanism

which transmission is continuously delayed as shown in Fig 1, a queue is created to transmit Node C and the other nodes that transmit with C. So, several queues can be partially made in this research, two queue were used.

This distributed precedence queue protocol provides the opportunity to create precedence queues for all nodes in a network. And, in the case that several precedence queues exist, each precedence queue assigned a priority so that they can be implemented independently.

The DPQ mode ID, which is stored in the 11 bit standard ID field shown in the Fig 4, indicates the precedence queue order of each node. Whenever a node carries out a transmission, it moves to the end of the queue, thus lowering its precedence to the minimum. All of the nodes following the transmitting node advance by one position in the queue, occupying the space that has just been created. Using this round-robin policy, collisions among messages are avoided.

The queue is not stored in some specific location. Instead, it is distributed among all the nodes in the network. Each node is responsible for storing and updating. That is, if the maximum permission delay time is reached, it creates a precedence queue, and then it has to dynamically change priorities to transmit preferentially with other nodes. And a precedence queue has to be dissolved when is completed an urgent task.

We suppose a network that is composed of Nodes A to G as shown in Fig 2. If Node C builds up a queue, the ID that is entered into the data frame queue can transmit and designate to 7 by lower 7 byte. At this time, it will be designated precedence priority to higher byte. Then, each node filters to enter itself into the queue, and it assigns its queue. After Node C transmits a message, it will go to the last position in the queue. And the other nodes will move up one position by order. And the remaining nodes that to be transmitted are designated using the upper 1 byte as shown in Fig 3; their queues will be dissolved or maintained using the upper 1 byte, as shown the Fig 3 after all transmissions are completed.

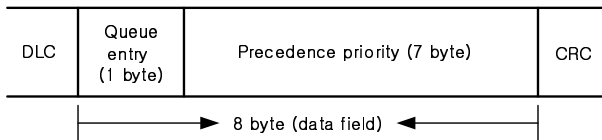


Fig. 3. Structure of a data field for DPQ

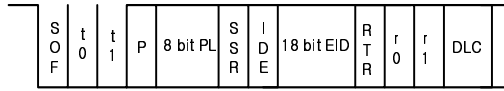
### 3.2 DPQ (Distributed Precedence Queue) Implementation Issues

The DPQ mechanism can be implemented without any modifications to the basic format of CAN frames. It uses an identifier field to designate the priority queue. Because the length of the conventional identifier field defined in the CAN standard is too small, the CAN extended format can be adopted.



Fig. 4. Format of the header of extended CAN frames

The DPQ uses the first 11 bits of the identifier field for its control information, whereas the remaining lower order 18 bits (ID ext.) are used to dynamically store the effective identifier of the an exchanged object (EID).



**Fig. 5.** Format of the header DPQ frames

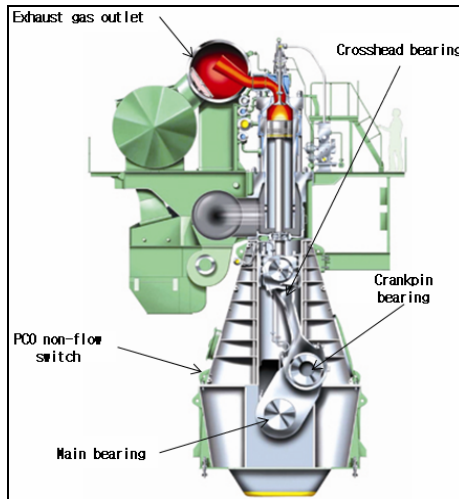
The first two bits ( $t_0$ ,  $t_1$ ) must be set at the logical value of zero as shown in Fig 5. Then, the protocol is divided by a standard CAN communication and DPQ mechanism. So, DPQ always has a higher priority than a CAN mechanism, and they can exist in this same space.

The priority bit  $P$  specifies whether the frame has to be transmitted as a high priority frame ( $P=0$ ) or as a low priority frame ( $p=1$ ). When  $T_1$  and  $P$  are used, the priority can be assigned a maximum 4 queues.

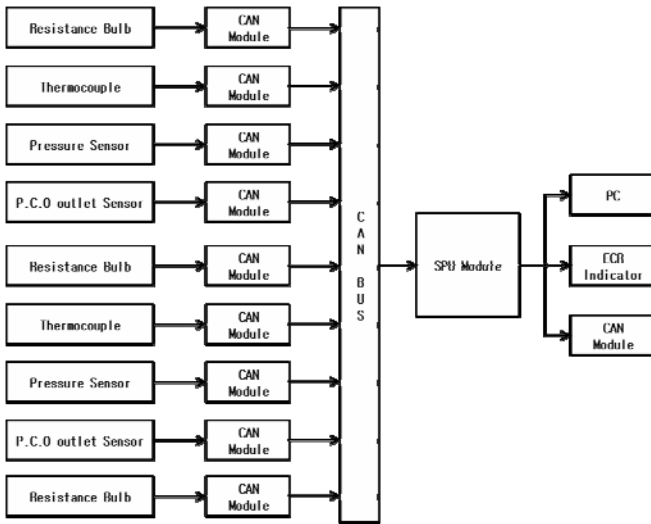
The next 8 bits represent the precedence level of the frame. Namely, these 8 bits show the transmission queue order. The DPQ, which was used in this research, uses  $t_0$ ,  $t_1$ , and then distinguishes the standard CAN mechanism, and sets each queue using  $P$ , and concludes the precedence in the queue using 8 bits.

## 4 System Architecture and Experiment

In this paper, we experiment with nine CAN communication Module and one SPU module by using sensor that the actual machine is used in Marine engines in order to evaluate the effectiveness of the proposed method.



**Fig. 6.** Marine Engine



**Fig. 7.** ESMS(Engine State Monitoring System) Architecture

Fig. 7 is whole structure of the Marine ESMS which proposes from the present paper. The basic principle of condition of the engine inside and the system which perceives a bearing attrition is crosshead header where is each cylinder of two-stroke diesel engines which are enormous cylinder stack gas temperature of engine inside, cooling fresh water exhaust an issue, Main bearing temperature and the sensor of the various type which perceives Marine engine piston cooling oil the perception flowing (PCO non-flow switch) switch conditions CAN communication modules leads and transmits in SPU (Signal Processing Unit). Data transmitted through the internal alarm and external alarm and sends a warning signal PC, ER(Engine Room), ECR(Engine Control Room) indication unit of such transfer is proposed to allow real-time monitoring.



**Fig. 8.** Hardware structure of CAN communication module

Fig. 8 is a CAN communication module. Each object was used dsPIC30F4012 with CAN communication module and MCP2551 was used to CAN transceiver.

SPU Module is consist of TMS320F28335 that combines a CAN Module and we used the SN65HVD230 for CAN transceiver. All of the transmission cycle of each object is set to 1ms, conflict was to occur frequently. Transmission message of the

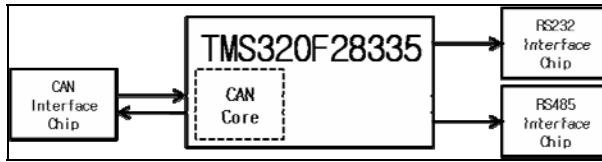


Fig. 9. Hardware structure of SPU module

object one has a highest priority and progressively lower priority. Communication speed was set at 1Mbps and object 1, 4, 7 was applied to the DPQ technique. The data of sensor for the bearing wear detection were managed. We can checking the Processed data in real time at SPU by RS232C communication at the PC and data are indicated in ER(Engine Room), ECR(Engine Control Room) by RS485 communication.

Table 1. Identification (ID) Definition

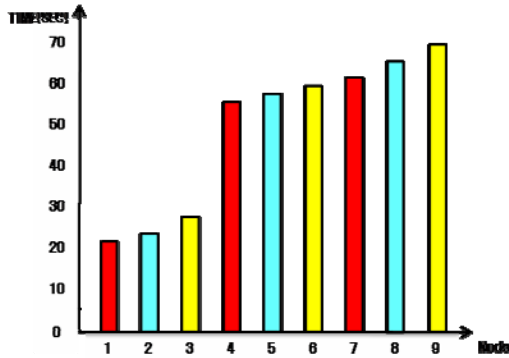
Node	Identification(ID, 11bit)	
	Standard CAN	DPQ
1	11 1 0000 0001	00 0 0000 0001
2	11 1 0000 0010	11 1 0000 0010
3	11 1 0000 0011	11 1 0000 0011
4	11 1 0000 0100	00 0 0000 0010
5	11 1 0000 0101	11 1 0000 0101
6	11 1 0000 0110	11 1 0000 0110
7	11 1 0000 0111	00 0 0000 0011
8	11 1 0000 1000	11 1 0000 1000
9	11 1 0000 1001	11 1 0000 1001

## 5 Result and Analysis

Figure. 8 represents substantial components of ESMS(Engine State Monitoring System). Power Supply Module and CAN Connector is in the left side, CAN Module is in the middle, and SPU is the right side. Sensor values in the each CAN Module which equipped with Resistance Bulb, Thermocouple, Pressure Sensor and P.C.O outlet sensor are converted through AC Conversion. And transmitted to SPU(Signal Processing Unit) through CAN bus. Note that, Node1 is SPU which integrates data from CAN Module, Node2 is Resistance Bulb which detects high temperature from Exhaust gas outlet, Node3 is Thermocouple, and Node4 is Pressure Sensor. Assignment of priority is importance of operating factors in the engine. So, high temperature node gets higher priority, and low temperature node gets lower priority.

Table 2. shows the completed time of 10000 times transmission which applied Standard CAN.

**Table 2.** Standard CAN Transmission Timing



**Table 3.** DPQ Mechanism Transmission Timing

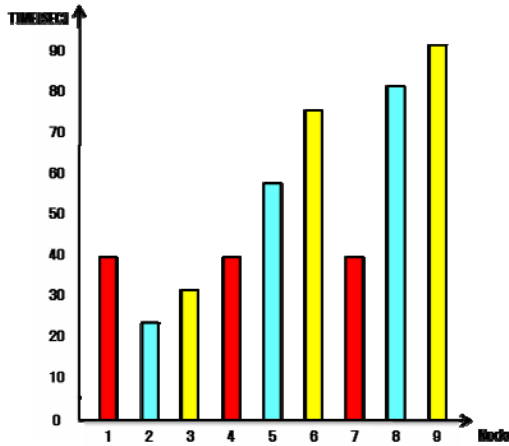


Table 3. shows the completed time of 10000 times transmission when DPQ Algorithm was applied to Node 1,4,7.

## 6 Conclusion

In this paper, the Marine engine state monitoring System using DPQ mechanism is developed to improve efficiently of priority which arbitrates collision in networks using CAN that is suitable for real time and distribution processing. Through these methods, lower Priority object 4 and 7 ensure transmission opportunity and delay time is improved. Features of the Marine engine monitoring system are as follows.

1. CAN communication module is connected to sensor used in actual Marine, is distinguished according to set ID and is designed for adding extra modules easily.



2. SPU(Signal Processing Unit) module for processing total data which transmit in CAN module is able to check Whole engine, bearing state and etc.
3. Whole states of Marine engine is able to be checked through the output values (voltage, current, resistor) of all sensors using the SPU (Signal Processing Unit) module.
4. Conventional communication method as RS232 and Rs485 is supported for compatibility with the existing Marine engine system. Therefore, state of Marine is able to be checked in ER (Engine Room) and ECR (Engine Control Room).

**Acknowledgments.** This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the Human resource development program for robotics support program supervised by the NIPA(National IT Industry Promotion Agency). (NIPA-2010-(C7000-1001-0009))

## References

1. Bosch: CAN Specification Version 2.0., Robert Bosch GmbH. Stuttgart (1991)
2. International Standard Organization, Road - vehicles Interchange of digital information - Controller area network for high-speed communication, ISO 11898 (November 1993)
3. Cena, G., Valenzano, A.: An Improved CAN Fieldbus for Industrial Application. IEEE Trans. on Industrial Electronics 44(4) (August 1997)
4. Hilmer, H., Kochs, H.D., Dittmar, E.: A Fault-Tolerant Communication Architecture for Real-Time Control System. In: IEEE International Workshop on Factory Communication System, pp. 111–118 (1997)
5. Corriganm, S.: Intriduction to the Controller Area Network, TI, Application Report, Texas (2002)
6. Farsi, M., Ratcliff, K., Barbosa, M.: An Overview of Controller Area Network. Computing and Control Engineering Journal 10(3), 113–120 (1999)
7. Guerreoro, C.: Hardware Support for Fault Tolerance in Triple Redundant CAN Controllers. In: 9th International Conference on Electronics, Circuits and System, vol. 2, pp. 457–460 (2002)
8. Guerreoro, C.: Design and Implementation of a Redundancy Manager for Triple Redundant CAN Controllers. In: IEEE 28th Annual Conference of the Industrial Electronics Society, vol. 3, p. 2294–2299 (2002)
9. IEEE Standards for Local Area Networks, Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications, ANSI/IEEE Std 802.3- ISO/DIS 8802/3 (1985)
10. Tindell, K., Burns, A.: Guaranteeing Message Latencies on Control Area Network (CAN). In: Proc. 1st International CAN Conference, Mainz, Germany (September 1994)

# Modeling and Fault Tolerant Controller Design for PM Spherical Actuator

Zheng Li

School of Electrical Engineering and Information Science, Hebei University of Science and Technology, Shijiazhuang, 050018, China  
lzhfgd@163.com

**Abstract.** This paper presents a method based on neural networks for achieving fault tolerant control in the PM spherical actuator control scheme. The model of a PM spherical actuator was developed. Tuning rules of the RBF networks which guarantees the stability of the fault system were derived and the on-line fault tolerant control scheme was developed. The control scheme does not need fault detection and diagnosis modules. An application tracking control problem for the tracking errors and Euler angles of an actuator driven by independent stator and rotor pairs is solved by using the controller. The effectiveness of the proposed method is illustrated by performing the simulation of a space trajectory tracking control.

**Keywords:** Fault tolerant; Neural networks; Spherical actuator; Controller design.

## 1 Introduction

The optimal control, fuzzy control and neural network control methods have been applied to control various autonomous robotic devices [1]-[6]. In particular, the fuzzy control and the neural-network control are attractive as practical robot control strategies because no mathematical models are needed and it can easily be applied to nonlinear and linear systems. It should be noted, however, that when designing a fuzzy controller, certain controller parameters must be tuned by trial and error. Such parameters include gains for its input data and output data, the number of membership functions, the width of a single membership function, and the number of control rules. Thus, it is important for control engineers to automate the parameter tuning of the fuzzy controller.

In recent years, the spherical motors with 3 degrees-of-freedom have attracted special interest as novel direct drive type actuators for many modern devices applications, such as robotic joints. The spherical stepper motor as one of them can provide advantageous features over traditional drive mechanisms which are usually constructed by several conventional drive motors or actuators, each having one degree of freedom and reducing the position accuracy, stiffness, dynamic performance and efficiency of the system.

The goal of this paper is to present a passive fault tolerant control method for actuators or robots of nonlinear system using RBF networks. To some nonlinear objects,

it is usually to derive linear model by linearization after modeling among equilibrium points and thus the control strategy can be adopted and incorporated. Due to the facts that some parts or elements are in fault state, the performance of the system can not achieve to a desired state. The application of neural networks into the fault tolerant control to overcome the nonlinear problems has been paid more attention, especially for the online self-study modes [7]-[12]. In the RBF NN designs, the weights of the neural networks are updated from the steady theory, so the dynamic change of the system can be quickly captured insuring the fault stability of whole system.

## 2 Control Method

### 2.1 Control Problem Presentation

Considering a fault system follow-up control with forms as follows [1]:

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{F}(\mathbf{x}) + \mathbf{B}_0 \mathbf{u} \quad . \\ \mathbf{y} &= \mathbf{C}\mathbf{x} \end{aligned} \tag{1}$$

where  $x \in R^{n \times l}$  denotes the state vector,  $u \in R^{n \times l}$  denotes the control vector,  $C$  is the chosen  $p$  states to be tracked, the other  $n - p$  states are in a certain field.  $y \in R^{p \times l}$  is the output vector.  $B_0$  denotes the input matrix in normal state of the system, in abnormal state is  $B_f$ , the nonlinear function  $F(\bullet)$  is smoothly continuous. The rank of  $(C \times B_0)$  is  $p$ , and also in fault state  $(C \times B_f)$  is  $p$ . In this way, the continuous  $u_d(t)$  starts from  $x_0$ , even in fault state of actuators the system can also track the desired output trajectory  $y_d(t)$  with its desired state of  $x_d$ ,

$$u_d = (CB_f)^{-1}(y_d - C^T F(x_d)) \quad . \tag{2}$$

The structure of the RBF NN fault tolerant control can be shown in figure 1. The output of the RBF neural network is expressed as

$$g(x) = \sum_{k=1}^h w_k \varphi_k(x) = W^T \varphi \quad . \tag{3}$$

where  $\varphi \in R^{h \times l}$  is Gauss function vector, and

$$\varphi_k(x) = \exp\left(-\frac{1}{\sigma_k^2} \|x - \mu_k\|^2\right), \quad k = 1, \dots, h \quad . \tag{4}$$

where  $h$  denotes the number of the neural elements,  $\sigma_k$  is the width of the Gauss function,  $\mu_k$  is the center of the Gauss function,  $w_k$  is the connected weights.

The system can be approached by a RBF neural network with enough neural elements within enough small errors, as

$$u_d(t) = \sum_{k=1}^h w_k^* \phi_k^* + \varepsilon(t) = W^{*T} \phi^* + \varepsilon \quad . \tag{5}$$

where  $w^* \phi^*$  is the optimal output of the RBF neural network. Due to the system is bounded, if the approached error  $\varepsilon(t)$  is bounded with its upper bound  $\varepsilon_h$ ,

$$\varepsilon_h = \sup_{t \in R^+} \|\varepsilon(t)\| . \tag{6}$$

Let  $e = x - x_d$ ,  $\dot{e} = \dot{x} - \dot{x}_d$ , the dynamic error of the fault system can be written as

$$\dot{e} = \dot{x} - \dot{x}_d = F(x) + B_f u - F(x_d) + B_f u_d . \tag{7}$$

Linearization of the above equation in the area of the  $x_d$  :

$$F(x) = F(x_d) + A(x - x_d) . \tag{8}$$

where  $A = \left. \frac{\partial F(x)}{\partial x^T} \right|_{x=x_d}$ . From figure 1, it can be seen

$$u = \hat{W}^T \hat{\phi} + K_p e . \tag{9}$$

where  $\hat{W}^T \hat{\phi}$  denotes the real output of the RBF neural network,  $K_p e$  is the  $P$  controller.

Substitute equation (5) and (9) to (7), the dynamic error equation of fault system can be shown as

$$\dot{e} = (A + B_f K_p) e + B_f (\hat{W}^T \hat{\phi} - W^{*T} \phi^* - \varepsilon) . \tag{10}$$

To be more simply, let  $J = A + B_f K_p$ ,  $\tilde{W} = W^* - \hat{W}$ ,  $\tilde{\phi} = \phi^* - \hat{\phi}$ , and

$$W^{*T} \phi^* = (\tilde{W}^T + \hat{W}^T)(\tilde{\phi} + \hat{\phi}) \tag{11}$$

thus  $\dot{e} = J e + B_f (\tilde{W}^T \hat{\phi} - W^{*T} \phi^* - \varepsilon)$ .

### 2.2 Online Rules Regulation

We choose Lyapunov function

$$V = \frac{1}{2} e^T P e + \frac{1}{2} \text{tr}(\tilde{W}^T \theta \tilde{W}) + \frac{1}{2} \tilde{\phi}^T \Lambda \tilde{\phi} \tag{12}$$

where  $P \in R^{n \times n}$  is definite matrix,  $\theta, \Lambda \in R^{k \times k}$  denotes the non-negative definite matrix with its derivative

$$\dot{V} = \frac{1}{2} [e^T \dot{P} e + e^T P \dot{e}] + \text{tr}(\tilde{W}^T \dot{\theta} \tilde{W}) + \tilde{\phi}^T \Lambda \dot{\tilde{\phi}} \tag{13}$$

Substitute (10) to (13) yields

$$\dot{V} = e^T Q e - \varepsilon^T B_f^T P e - \tilde{\phi}^T \hat{W} B_f^T P e - \hat{\phi}^T \tilde{W} B_f^T P e + \text{tr}(\tilde{W}^T \dot{\theta} \tilde{W}) + \tilde{\phi}^T \Lambda \dot{\tilde{\phi}} \tag{14}$$

which can find the only solution, where  $Q = \frac{1}{2}(J^T P + P J)$

However 
$$\text{tr}(\tilde{W}^T \dot{\theta} \tilde{W}) = \sum_{i=1}^p \tilde{W}_i^T \dot{\theta} \tilde{W}_i \tag{15}$$

$$\hat{\phi}^T \tilde{W} B_f^T P e = \sum_{i=1}^p \hat{\phi}^T \tilde{W}_i B_{fi}^T P e \tag{16}$$

where  $\tilde{W}_i$  is the  $i$  th column of matrix  $\tilde{W}$ ,  $B_{fi}$  is the  $i$  th column of matrix  $B_f$ , the equation turns to

$$\dot{V} = -\dot{e}^T Q e - \varepsilon^T B_f^T P e + \tilde{\phi}^T (-\hat{W} B_f^T P e + \Lambda \tilde{\phi}) + \sum_{i=1}^p (-\tilde{W}_i^T \hat{\phi} B_{fi}^T P e + \tilde{W}_i^T \theta \tilde{W}_i) \tag{17}$$

To choose the following weights and regulation rules of  $\phi$

$$\tilde{W} = \theta^{-1} \hat{\phi} B_f^T P e \quad i = 1, \dots, p \tag{18}$$

$$\tilde{\phi} = \Lambda^{-1} \hat{W} B_f^T P e \tag{19}$$

and  $\dot{V} = \dot{e}^T Q e - \varepsilon^T B_f^T P e$ , substitute to (6) with  $\dot{V}$  is negative definite,

$$\dot{V} \leq -\|e\|_2 \left[ \lambda_{\min}(Q) \|e\|_2 + \|P\| \|B_f\| \|\varepsilon_h\| \right] \tag{20}$$

Let  $\|B_f\| \|\varepsilon_h\| = \delta_{eh}$ , when  $\|e\|_2 \geq \frac{\|P\|}{\lambda_{\min}(Q)} \delta_{eh} = E$ ,  $\dot{V}$  is negative definite.

Since the unknown values of  $B_f$ , it can be expressed as  $B_f = B_0 + \Delta B$  and can be disposed approximately as  $B_f = B_0$ .

The regulation rules thus can be further changed to

$$\tilde{W}_i = \theta^{-1} \hat{\phi} B_{0i}^T P e, \quad i = 1, \dots, p \tag{21}$$

$$\tilde{\phi} = \Lambda^{-1} \hat{W} B_{0i}^T P e \tag{22}$$

Due to  $\phi_k = g(\mu_k, \sigma_k)$ , the center  $\mu_k$  of RBF Gauss function  $\phi_k$  is derived by learning algorithm, and the width of RBF Gauss function  $\phi_k$  is derived by equation (22) [1].

From equation (17), (21) and (22), it can be analyzed that

$$\dot{V} = -e^T Q e - e^T P (B_0 + \Delta B) \varepsilon - (\tilde{\phi}^T \hat{W} \Delta B P e + \hat{\phi}^T \tilde{W} \Delta B P e) \tag{23}$$

when  $\|\Delta B\| = 0$ , it equals to (19). If  $\|\Delta B\| \leq M$ , let

$$\left\| (\tilde{\phi}^T \hat{W} + \tilde{\phi}^T \tilde{W}) \right\| \|P\| = \zeta(\tilde{\phi}, \tilde{W}) \tag{24}$$

And when  $\|e\|_2 \geq \frac{\|P\|}{\lambda_{\min}(Q)} \delta_{eh} + \frac{\zeta M}{\lambda_{\min}(Q)} = E + \Delta E$  (25)

$\dot{V}$  is negative definite. The last term of above equation is caused by the actuator or element faults, and enlarged the error. It also can derive better fault tolerant control

effects in the case of  $\tilde{\phi}$  and  $\tilde{W}$  with small values by properly adjusting the parameters or the error is restricted in a bounded area.

So if equation (25) can be satisfied,  $V$  is negative definite and the system is stable whenever normal or fault state to realize the real-time fault tolerant control.

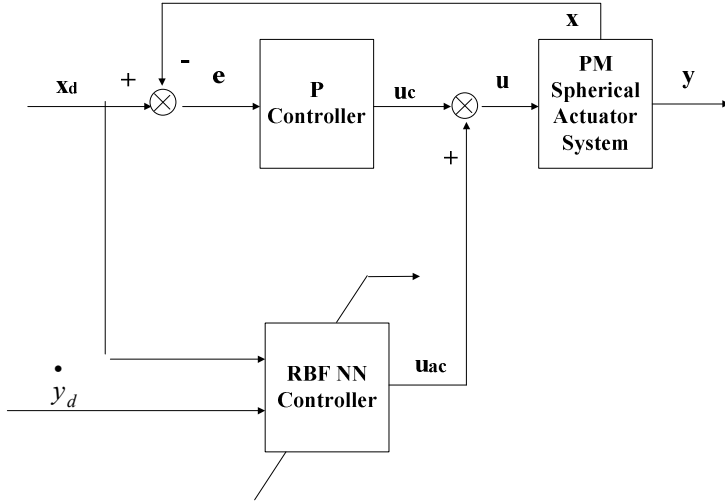


Fig. 1. Fault tolerant control scheme based on RBF neural network.

### 2.3 Modeling of the PM Spherical Actuator

The motor prototype consists of a ball-shaped rotor with four layers of 40 poles and two layers of 24 poles. There are 10 poles with equally spaced position of alternative N-S distribution for the rotor. There are 12 poles in each layer and controlled separately for the stator. Fig. 1 shows the basic structure of the actuator. The actuator is designed to implement maximum 67 degrees tilt motion and 360 degrees rotation.

Using the Lagrange energy method, the dynamic equations of spherical actuator are obtained. Each equation represents an axis of Cardian Angle. It can be seen that there are many one-order and two-order coupling terms in the equation, which indicates serious inter-axes nonlinear coupling of the spherical actuator.

Like the conventional reluctance motors, the operation of the spherical stepper motor is also based on the reluctance forces. The difference is that the flux path is not closed due to the non-magnetic material used in the rotor. It can be considered that the torque or force exerted on the rotor is the summation of individual rotor/stator pair interactions and the torque calculation model has the linear property.

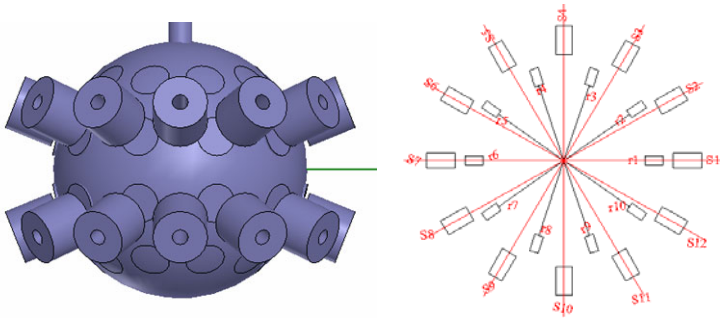


Fig. 2. Basic structure of the actuator

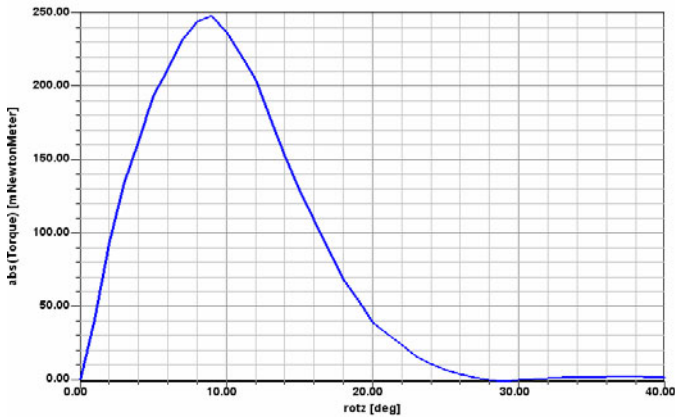
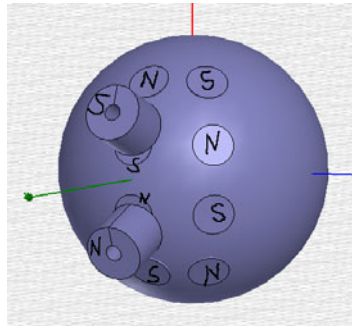


Fig. 3. The illustration of the torque produced by one pair of poles

The torque equation can be combined into one vector equation in the x-y-z coordinate [13-16]:

$$\mathbf{T} = \frac{1}{2} \sum_{j=1}^{n_r} \sum_{k=1}^{n_s} \frac{\partial \Lambda_{jk}}{\partial \delta_{jk}} F_{jk}^2 \frac{\mathbf{r}_{rj} \times \mathbf{r}_{sk}}{\|\mathbf{r}_{rj} \times \mathbf{r}_{sk}\|} \tag{26}$$

From equation (1), the torque calculation has been simplified to the superposition of torque component by individual pair. The former two parameters can be taken as torque constant, and derived by accurate computation or experiments. The torque produced by single stator/rotor pair can be written as

$$\mathbf{T}_{jk} = k(\delta_{jk}, i_k) \frac{\mathbf{r}_{rj} \times \mathbf{r}_{sk}}{\|\mathbf{r}_{rj} \times \mathbf{r}_{sk}\|} Ni_j \tag{27}$$

where the constant  $k(\delta_{jk}, i_k) = \frac{\|\mathbf{T}_{jk}\|}{Ni_j}$  is a function of  $\delta_{jk}$  and  $i_k$ , so in the case of the reference torque is known, the control current corresponding to some rotor orientation can be solved by

$$\mathbf{T} = \mathbf{K}_T \mathbf{I}_d \tag{28}$$

$$\mathbf{I}_d = \mathbf{K}_T^+ \mathbf{T} \tag{29}$$

where  $\mathbf{K}_T$  is the torque constant matrix about some orientation,  $\mathbf{K}_T^+$  denotes its general inverse matrix.  $\mathbf{I}_d$  is the reference control current.

By development and substitution, the complete mathematical model of the system can be written as

$$\mathbf{M}(\mathbf{Q})\ddot{\mathbf{Q}} + \mathbf{C}(\mathbf{Q}, \dot{\mathbf{Q}})\dot{\mathbf{Q}} + \mathbf{G}(\mathbf{Q}) + \mathbf{T}_{Qf} = \mathbf{T}_{Qe} \tag{30}$$

where

$$\mathbf{M}(\mathbf{Q}) = \mathbf{J}_Q^T \mathbf{M}_r \mathbf{J}_Q = \begin{pmatrix} J_{bac}\beta^2 + J_b s\beta^2 & 0 & J_b s\beta \\ 0 & J_{ba} & 0 \\ J_b s\beta & 0 & J_b \end{pmatrix},$$

$$\mathbf{C}(\mathbf{Q}, \dot{\mathbf{Q}}) = \mathbf{J}_Q^T \mathbf{M}_r \dot{\mathbf{J}}_Q + \mathbf{J}_Q^T \mathbf{C}_r \mathbf{J}_Q,$$

$$\mathbf{G}(\mathbf{Q}) = \mathbf{J}_Q^T \mathbf{T}_g = m_c g r_c \begin{pmatrix} -s\alpha c\beta^2 c\gamma + s\beta c\beta s\gamma \\ -s\alpha c\beta s\gamma - s\beta c\gamma \\ 0 \end{pmatrix},$$

$$\mathbf{T}_{Qf} = \mathbf{J}_Q^T \mathbf{T}_f,$$

$$\mathbf{T}_{Qe} = \mathbf{J}_Q^T \mathbf{T}_e$$

The nonlinear friction has a great effect on the servo system and is hard to derive accurate mathematical models. Here, the robust design is performed by confirming its upper bound  $\rho_f$ , and was estimated by

$$\mathbf{T}_{Qf}(\dot{\mathbf{Q}}) = \text{sign}(\dot{\mathbf{Q}})\mathbf{T}_c + \text{sign}(\dot{\mathbf{Q}})(\mathbf{T}_s - \mathbf{T}_c) \exp(-\mu|\dot{\mathbf{Q}}|) \tag{31}$$

where  $\mathbf{T}_s$ ,  $\mathbf{T}_c$ ,  $\mu$  denote the static friction, coulomb friction and exponential constant.

The motor dynamics have the same form and properties as those of a robotic manipulator. The robust control scheme can be applied on the continuous trajectory



tracking control, to minimize the effects of uncertainties including parameters perturbation, model error and outer disturbance.

### 3 Simulation Examples

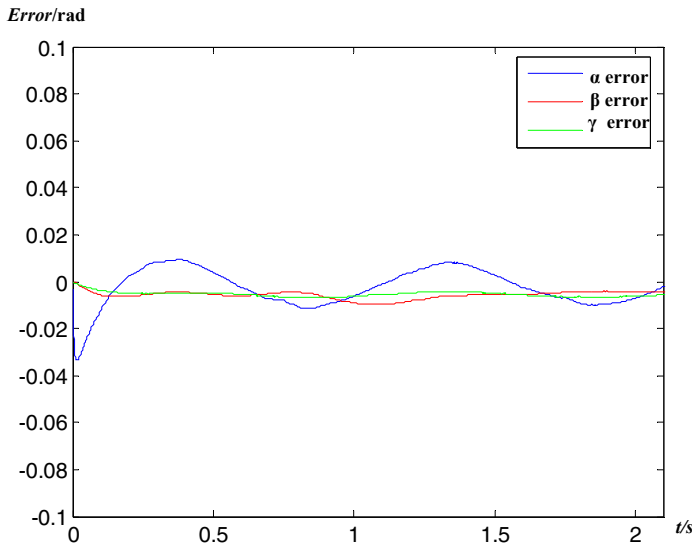
Simulation has been carried out with the presented robust neural network control algorithm and the system model. Table 1 shows the system parameters, the configuration parameters can be seen from reference. The desired continuous trajectory is

$$(Q_d = [\sin \pi, \cos 2\pi, 0.6(1 - e^{-5t^2})]) \tag{32}$$

with the sinusoidal, exponential, and linear relations. Considering a random disturbance torque with amplitude of 2.5Nm is exerted on the rotor, the simulation is to prove the validity of the control strategy. The results validate that in the case of the fault the whole system is stable with a satisfied tracking performance and its fault tolerant ability is special in integrality as shown in figure 4 to figure 7.

**Table 1.** System added parameters

Parameters	Values
Torque Model Fitting order $M, N$	6, 3
Load mass $m_c$	0.25 kg
Distance to the centroid $r_c$	0.2m
Friction model $T_c, T_s, \mu$	0.4, 0.7, 6



**Fig. 4.** Curves of tracking errors convergence

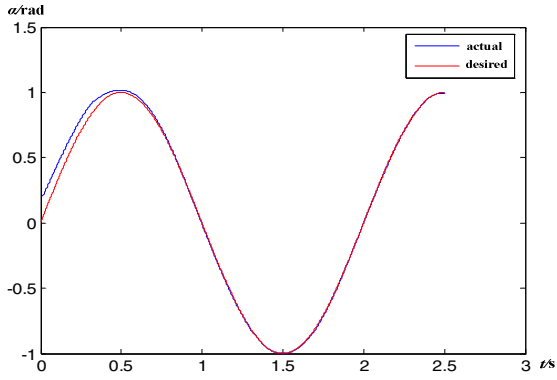


Fig. 5. Curves of  $\alpha$  tracking process

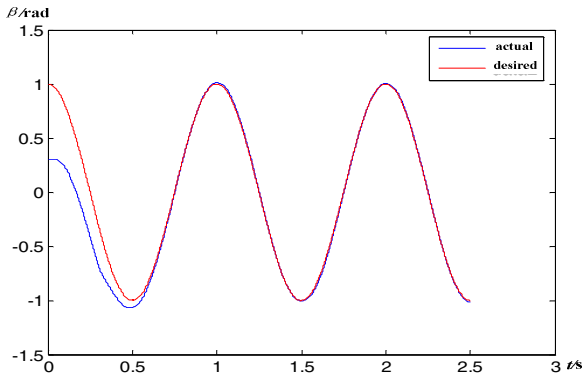


Fig. 6. Curves of  $\beta$  tracking process

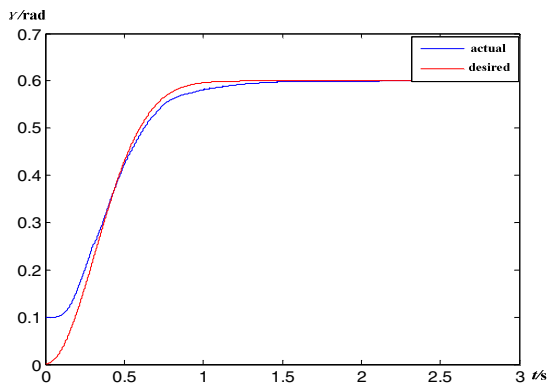


Fig. 7. Curves of  $\gamma$  tracking process

## 4 Conclusions

A tracking control problem for the PM spherical actuator driven by independent stator/rotor poles has been solved by using RBF neural network fault tolerant controller. The tuning rules of the RBF networks were derived and guarantee the stability of the fault system developing an on-line fault tolerant control scheme. The simulation results of a trajectory tracking control with fault conditions show the effectiveness of the proposed control scheme. It is expected the results can give basic references for further spherical actuators and fault tolerant control researches.

**Acknowledgments.** This work was supported in part by National Natural Science Foundation of China under Grant no. 50677013 and in part by Natural Science Foundation of Hebei Province of China under Grant no. E2009000703, and by a grant from the National High Technology Research and Development Program of China (863 Program) (No. 2007AA04Z214).

## References

1. Gomi, H., Kawato, M.: Neural networks control for a closed-loop system using feedback error learning. *Neural Networks* 7(6), 933–946 (1993)
2. Katebi, M.R., Grimble, M.J.: Integrated control, guidance and diagnosis for reconfigurable autonomous underwater vehicle control. *International Journal of Systems Science* 30(1), 1021–1032 (1999)
3. Saluds, S., Fuente, M.J.: Neural-networks-based fault detection and accommodation in a chemical reactor. In: *Proc. of 14th World Congress of IFAC*, pp. 169–174. IEEE Press, Beijing (1999)
4. Katebi, M.R.: Guidance controller design for autonomous underwater vehicles. In: *Proc. of 14th World Congress of IFAC, Beijing*, pp. 191–196 (October 1999)
5. Kosko, B.: *Neural Networks and Fuzzy Systems*. Prentice-Hall, Englewood Cliffs (1992)
6. White, D.A., Sofge, D.A.: *Handbook of Intelligent Control-Neural, Fuzzy, and Adaptive Approaches*. Van Nostrand Reinhold, New York (1992)
7. Horikawa, S., Furuhashi, T., Uchikawa, Y.: On fuzzy modeling using fuzzy neural networks with the backpropagation algorithm. *IEEE Trans. Neural Networks* 3(5), 801–806 (1992)
8. Hasegawa, T., Horikawa, S., Furuhashi, T., Uchikawa, Y.: An application of fuzzy neural networks to design of adaptive fuzzy controllers. In: *Proc. Int. Joint Conf. Neural Networks*, pp. 1761–1764. IEEE Press, New York (1993)
9. Moody, J., Darken, C.J.: Fast learning in networks of locally-tuned processing units. *Neural Computa.* 1, 281–294 (1989)
10. Jang, J.-S.R., Sun, C.-T.: Functional equivalence between radial basis function networks and fuzzy inference systems. *IEEE Trans. Neural Networks* 4(1), 156–159 (1993)
11. Nie, J., Linkens, D.A.: Learning control using fuzzified selforganizing radial basis function network. *IEEE Trans. Fuzzy Syst.* 1(4), 280–287 (1993)
12. Watanabe, K., Tang, J., Nakamura, M., et al.: A Fuzzy-Gaussian Neural Network and Its Application to Mobile Robot Control. *IEEE Trans. on Control Systems Technology* 4(2), 193–199 (1996)

13. Wang, Q., Li, Z., et al.: Magnetic Field Computation of a PM Spherical Stepper Motor Using Integral Equation Method. *IEEE Trans. on Magn.* 42(4), 731–734 (2006)
14. Qunjing, W., Zheng, L., Kun, X., et al.: Calculation and Analysis on Configuration Parameters and Torque Characteristics of a Novel Spherical Stepper Motor. *Proceedings of the CSEE* 26(10), 158–165 (2006)
15. Qunjing, W., Zheng, L., et al.: Kinematic Analysis and Simulation of Permanent Magnet Spherical Stepper Motor. *Journal of System Simulation* 17(9), 2260–2264 (2005)
16. Li, Z., Wang, Q.: Modeling and Control of a Permanent Magnet Spherical Stepper Motor. In: *Proceeding of International Conference on Electrical Machines and Systems*, pp. 1574–1579. IEEE Press, Seoul (2007)

# Artificial Neural Network to Predict the Surface Maximum Settlement by Shield Tunneling\*

Jinli Qiao<sup>1</sup>, Jianqin Liu<sup>2,\*</sup>, Wei Guo<sup>2</sup>, and Yitong Zhang<sup>2</sup>

<sup>1</sup> School of Civil Engineering, Hebei University of Technology, Tianjin 300132, China

<sup>2</sup> School of Mechanical Engineering, Tianjin University, Tianjin 300072, China

Tel.: +86 013516187323; Fax: +86 022 27404559

liujianqin@tju.edu.cn

**Abstract.** Numerous empirical and analytical relations exist between shield tunnel characteristics and surface deformation. Artificial neural networks (ANN) was used to develop predictive relations between the maximum surface settlement and shield tunnel overburdens, shield diameters, thrusts of shield tunneling, advancement rates of shield, fill factors of grouting, cohesive forces, friction angles and compression modules of the soils. So, ANN can become a useful predictive method. With the advantage of ANN in nonlinear problem, the theoretical model to predict the maximum surface settlement is established. The agreement of the measured results with the actual situation of being predicted shows that the proposed model is satisfactory.

**Keywords:** artificial neural network; the maximum surface settlement; shield tunneling.

## 1 Introduction

Due to recent city developments with limited available land to build on, more and more public facilities are developed under the ground surface. Construction of shield tunnel excavation process in poor self-formation soils induces generally soil movement, which could seriously affect the stability and integrity of existing structures (pile foundations, buildings, et al). With this tunneling technique, ground movement can be, in theory, controlled by balancing the pressure inside the earth pressure chamber relative to the outside ground pressure during excavation.

To control the ground movement in the design and construction of shield tunnel, we must understand the law of ground movement as much as possible to accurately predict settlement, the settlement range, and analyze the various factors affecting the settlement, but whether it is the internal deformation of the tunnel, or surface deformation corresponding to the top of the tunnel, is a complex nonlinear dynamic systems, and it is difficult to reveal its inherent laws using traditional methods and techniques.

---

\* Project 50975194 supported by National Natural Science Foundation of China.

Project 2007CB714000 supported by National Basic Research Program of China.

\*\* Corresponding author.

Surface settlement caused by shield construction is analyzed using the most construction experience and the settlement slot curve proposed by Peck. Peck [1] basing on statistical analysis of field data of mountain tunnels, artificial tunneling shield machine, and semi-mechanized shield machine, in addition to pneumatic shield construction machines and other engineering methods proposed that the shape of surface settlement trough shield construction is similar to the normal distribution curve in probability theory, and made the settlement formula of various points on the surface settlement tank.

$$\delta(x) = \delta_{\max} \cdot \exp\left(-\frac{x^2}{2i^2}\right) \quad (1)$$

Where:  $\delta(x)$  is the settlement of the surface point  $x$  away from the center;  $i$  is the width factor of settlement trough (the distance of tunnel center and settlement curve inflection point);  $\delta_{\max}$  is the center's maximum settlement  $\delta_{\max} = \frac{\nu A_c}{2.5i}$ ;  $A_c$  is the area of settlement trough,  $A$  is the cross-sectional area of tunnel;  $\nu$  is the loose factor of soil, the general value of 1% ~ 3%.

HS Lang et al measured the ground settlement of the sewage pipes of Taipei City Shield Construction, after the draw, the vast majority of the settlement occurred in the first 4 days after the shield being passed, while the final shape of the settlement trough is similar to Peck curve. In later studies, he has proposed that the curve of longitudinal settlement of earth pressure balance shield changing over time was hyperbolic type[2].

$$S_{\max} = \frac{t}{a + bt} \quad (2)$$

Where:  $a$ ,  $b$  is parameters related with soil properties shield test being obtained;  $t$  is the elapsed time. Surface settlement produced by tunnel construction is affected by many factors; all of this makes the shield-ground interaction complex, mainly include: tunnel overburdens, shield diameters, thrusts of shield tunneling, advancement rates of shield, fill factors of grouting, cohesive forces, friction angles and compression modules of the soils etc.

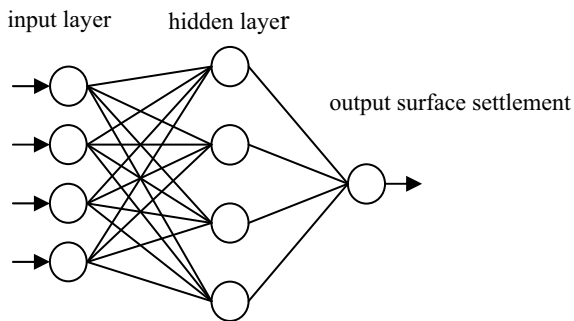
Empirical methods are still widely used; however, predictions of ground movements based on such methods are insufficient for most practical applications. There are a limited number of analytical and numerical tools that can be used to predict ground deformations, but there is a growing demand for developing practical rational methods for tunnel design. And many researcher have done a lot of theoretical analysis and testing for the surface settlement of shield tunnel and has been a lot of useful results[3-10] but because of the limited number of measurement and test data and the uncertainty of the relationship between various factors, it is very difficult to fully reflect the regular pattern on the ground settlement. The adjusted network is a good tool for settlement prediction of new tunnels in the same geological environment.

Furthermore, this analysis shows the influence of network training parameters and previous input data treatment on the quality of the adjustment obtained. Significantly enhanced neural network predictive ability was found due to the use of certain data processing techniques. Knowledge acquired was applied to further develop use of this technique for tunnel instrumentation.

## 2 Artificial Neural Networks

The structure of a neural network, in general, consists of an interconnected group of artificial neurons. These processing units receive the information, apply some simple processing on them and pass them to other neurons. The characteristics of a neural network come from the activation function and connection weights. Since the neural network stores data as patterns in a set of processing elements by adjusting the connection weights, practically, neural networks may be used in nonlinear statistical data modeling, system identification, extraction of complex relationships between inputs and outputs of a system, and for pattern recognition.

The structure of a simple neural network generally contains three layers: input layer, hidden layer, and output layer in Figure 1. Within the ANN training process, the number of hidden layers and the number of nodes in each layer depend on the complexity of the patterns and the nature of the problem to be solved. The training procedure consists of a sequential data feed into the network, followed by the comparative evaluation of the corresponding output provided by the ANN and the actual result. The network adjusts the weighting of the connection links in a continuous effort to produce the results that would best correspond to the training dataset. A complete pass of all the input data through the network consists a training epoch and usually a great number of epochs is required for the residual error to converge below a pre-specified threshold. The testing data set is used to verify the appropriateness of the trained network. Finally, output from other systems or models such as expert systems and statistics need to validate the results from ANN.



**Fig. 1.** Neural network architecture

ANN algorithm is summarized as follows:

- (1) Using a random number to chose weights and the initial value of thresholds.
- (2) The network input sample mode, input mode vector:

$$A_i(x_{i1}, x_{i2}, \dots, x_{ik}), i = 1, 2, \dots, m, \quad (3)$$

Where:  $m$  is the number of learning model,  $k$  is the number of the input layer neuron; the output vectors of the hope of the corresponding input mode,  $Y_n(y_1, y_2, \dots, y_n)$  is the output layer neuron number.

(3) The input mode of each unit of the hidden layer:

$$S_i = \sum_{j=1}^N \omega_{ij} \alpha_j - \theta_i, \quad j = 1, 2, \dots, N. \tag{4}$$

Where:  $\omega_{ij}$  is the right value of the input layer to the middle layer;  $\theta_i$  is the middle layer neuron threshold, N is the number of middle layer neurons.

(4) As the convergence of the common S-type function is slow, but with the increase in the number of iterations, the error steadily declines. Using improved BP algorithm, application of S-type function is appropriate, so this function  $S_i$  is used as the independent variables of S-type (Sigmoid) to calculate output of each hidden layer neuron, S-type function is the following function.

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{5}$$

To  $S_i$  generation into the  $f(x)$  :

$$a_i = f(S_i) = \frac{1}{1 + e^{-\sum_{j=1}^N \omega_{ij} \alpha_j - \theta_i}}. \tag{6}$$

(5) Input of each cell of the output layer and the network actual output.

$$b_k = f\left(\sum_{i=1}^N a_i \omega_{ik} - \theta_{bk}\right). \tag{7}$$

$$y = f\left(\sum_{i=1}^N b_k \omega_{ky} - \theta_y\right). \tag{8}$$

Where:  $\omega_{ik}$ ,  $\omega_{ky}$  are the connection weights of input layer and middle layer, middle layer and output layer respectively;  $\theta_{bk}$ ,  $\theta_y$  are the threshold of the middle unit and the output unit respectively; f is the S function.

(6) Calculate the error between the actual output value and hope output value, based on the size of the error adjusts the connection weights output layer to the underlying automatically, that is, propagation process of the errors  $\delta_y$  in the output layer to the error  $\delta_{bk}$  of the middle layer. However, the limitations of BP network itself, such as the existence of local minimum problems, slow convergence of learning algorithm, the number of the characteristics of each input sample must be the same and so on, in order to improve its limitations, in this paper, the standard correction formula is basically the introduction of weights momentum term to accelerate the learning rate, to prevent oscillation. Namely:

$$\Delta\omega(t+1) = \eta \frac{\partial E}{\partial \omega} + \alpha \Delta\omega(t). \tag{9}$$

Where:  $\alpha$  is the momentum factor, generally take 0.9 or so.



### 3 The ANN Model of the Surface Settlement Caused by Shield Excavation

#### 3.1 The Merits of ANN Using for the Surface Settlement

Artificial neural network model, in particular the BP ANN model is applied to the study of surface settlement successfully [11-13]. BP ANN model is superior to traditional methods in some ways, mainly reflected in:

(1) It is well known that surface settlement caused by shield excavation are complex nonlinear systems, because BP ANN model is essentially non-linear fitting, while it is generally better to take into account the non-linear characteristics of the surface settlement caused by shield excavation than the traditional log-linear model;

(2) We can see surface settlement caused by shield excavation has large uncertainty from the measured data, and the measured surface settlements have also a big difference even in the same condition. BP model can maximize to overcome this uncertainty comparing with the certainty of the formula of surface settlement;

(3) In order to make the established formula relatively simple, some minor parameters were removed and retained only the main ones in the traditional model of surface maximum settlement. BP model can regard any number of variables as input parameters, that is a more comprehensive for the parameters affecting the surface maximum settlement caused by shield excavation;

(4) The traditional formula for the surface maximum settlement is established by a limited number of laboratory data and the measured data, its use must be subject to certain limitations. The BP model is established through more information, thus its application is a broader scope. Although the BP model can not be used beyond their scope of learning and training, this problem can be resolved by expanding the scope of training samples.

(5) To the traditional model of surface maximum settlement, the user must have known the surface settlement theory caused by shield excavation, but just understanding the scope of information of BP model of the surface maximum settlement and without surface settlement theory, one can use BP model having established to predict the surface settlement.

#### 3.2 The Factors of the Surface Settlement in ANN Model

Select the nine main factors affecting the surface maximum settlement caused by the shield tunnel construction: 1)tunnel overburdens; 2)shield diameters; 3)thrusts of shield tunneling; 4) advancement rates of shield; 5)fill factors of grouting; 6) Grouting pressure; 7)cohesive forces; 8) friction angles; 9) Compression modules of the soils et al, and establish a network structure including the 9 input units and output units. According to data analysis, input units used: cohesive forces  $c$ , friction angles  $\varphi$ , compression modules  $E_s$  of the soil, tunnel overburdens  $H$ , shield diameters  $D$ , grouting pressure  $P$ , fill factors of grouting  $n$ , thrusts of shield tunneling  $F$ , advancement rates of shield. In the surface deformation caused by tunnel construction, the surface maximum settlement is the most concerned, so the output unit used: the surface maximum settlement  $S_{\max}$ .

### 3.3 The ANN Model of the Surface Maximum Settlement by Shield Tunneling

Generally, there is no direction and precise method to determine the most appropriate number of neurons for including in each hidden layer in the neural networks. This problem becomes more complicated as the number of hidden layers in the network increases. To establish an optimal network that can be used for predicting the surface maximum settlement, one needs to begin with training and testing the artificial neural networks using a subset of all data sets. In the pilot experiment data set, the samples are divided into a training set and a validation set. Networks with different numbers of hidden nodes will be trained all the way to the convergence of the training samples, measuring their performance with the validation set, and choosing the network that yields the best performance of the validation set. Finally, this selected network model will be used for the whole data set.

**Table 1.** Samples used for network training

No.	$C$ /KPa	$\varphi$ /°	$E_s$ /MPa	$H$ /m	$D$ /m	$P$ /MPa	$n$ /%	$F$ /MN	$v$ /mm·min <sup>-1</sup>	$S_{\max}$ /mm
1	10.0	25.0	9.12	20	6.34	0.34	1.4	10.00	20	35.1
2	11.4	19.2	8.42	12	6.4	0.4	1.4	12.00	30	42.4
3	32.5	15.5	7.67	9.4	6.34	0.3	1.5	15.00	20	40.6
4	214.9	23.8	34.2	18.2	6.34	0.4	1.7	20.00	30	10.5
5	12.9	11.8	7.22	15.5	6.4	0.25	1.5	13.00	40	52.4
6	15.0	13.7	8.21	12.0	6.34	0.45	1.5	14.00	20	40.5
7	12.0	16.6	4.24	11.8	6.25	0.25	1.6	16.00	30	79.6
8	11.8	15.2	5.91	9.78	6.34	0.55	1.7	14.00	20	47.3
9	11.7	16.4	6.48	8.4	6.4	0.3	1.7	14.00	30	53.3
10	32.4	10.7	11.17	14.5	6.40	0.25	1.7	31.65	60	22.0
11	31.2	42.1	35.7	15	6.4	0.30	1.5	30.00	30	14.1
12	11.9	13.8	5.22	11.9	6.34	0.40	2.0	14.00	20	21.2
13	12.1	13.7	5.21	12	6.34	0.35	1.7	14.00	30	55.3
14	201.7	23.5	35.2	20.6	6.40	0.30	1.2	30.00	20	11.2
15	36.7	20.7	7.26	13.8	6.4	0.25	1.7	31.65	60	7.5
16	32.4	12.4	11.17	14.2	6.4	0.25	1.7	31.65	60	14.8
17	43.6	30.0	9.12	14.5	6.4	0.25	1.5	31.65	60	8.9
18	34.2	9.20	7.26	14.5	6.4	0.25	1.7	31.65	60	22.8
19	34.2	14.5	7.26	14.5	6.4	0.25	1.7	31.65	50	16.4
20	340	44.9	35.02	12.0	6.25	0.2	0.9	33.00	30	16.8
21	240	30.1	23.12	24.0	6.25	0.3	1.2	33.00	30	19.2
22	12.0	13.4	5.01	12.7	6.34	0.4	2.0	14.00	20	32.8
23	11.8	14.4	5.53	10.9	6.34	0.4	2.0	14.00	20	27.1
24	11.9	13.8	5.22	11.9	6.34	0.4	1.8	14.00	20	38.6
25	11.9	13.8	5.22	11.2	6.34	0.3	1.7	14.00	30	89.9
26	11.9	13.8	5.22	10.5	6.34	0.25	1.4	14.00	40	62.5
27	12.0	13.7	5.21	12.0	6.34	0.35	1.5	14.00	30	57
28	12.0	13.6	5.2	11.8	6.34	0.35	1.4	14.00	40	79.6
29	11.8	15.2	5.91	9.78	6.34	0.35	1.7	14.00	20	47.3
30	28.1	32.8	5.42	13.3	6.40	0.25	1.5	31.65	40	9.6
31	12.2	13.1	5.2	12	6.4	0.35	2.0	14.00	20	20.3
32	12.0	13.8	5.21	11.8	6.34	0.4	1.7	31.65	30	45.1
33	34.0	16.6	7.26	12	6.4	0.3	1.6	16.00	20	70.2
34	112	35.2	25.9	9.78	6.34	0.35	1.7	31.00	20	17.3
35	30.0	16.5	7.60	10.4	6.25	0.3	1.6	16.00	40	35.4

**Table 2.** Testing sample for network

No.	$C$ /KPa	$\varphi$ /°	$E_s$ /MPa	$H$ /m	$D$ /m	$P$ /MPa	$n$ /%	$F$ /MN	$\nu$ /mm · min <sup>-1</sup>
1	11.2	19.5	8.26	6.1	6.34	0.3	1.4	14.00	40
2	11.8	14.7	5.67	10.4	6.34	0.4	1.8	14.00	20
3	11.9	14.1	5.35	11.4	6.34	0.4	1.8	14.00	20
4	30.2	22.8	11.8	20.0	6.25	0.25	1.2	23.00	40
5	15.3	23.9	3.87	12.0	6.4	0.3	1.5	31.65	40
6	18.7	13.5	6.2	10.6	6.34	0.30	1.6	14.00	30

**Table 3.** Comparison of measured and predicted values

No.	measured $S_{\max}$ /mm	predicted $S_{\max}$ /mm	Error
1	84.5	81.50	3.55%
2	40.6	41.60	2.46%
3	38.1	39.40	3.41%
4	34.0	36.10	6.18%
5	7.6	7.40	2.63%
6	26.4	25.45	3.59%

In this paper, the sample data shown in Table 1 are used as a pilot experiment (i.e., for training and validation) and thus for determining the optimal model. The data are fed into the ANN, where the input layer consist 9 input nodes that represent all influencing factors. The process attempts to establish the optimal neural network model and an appropriate number of training epochs for the problem. A validation set shown in

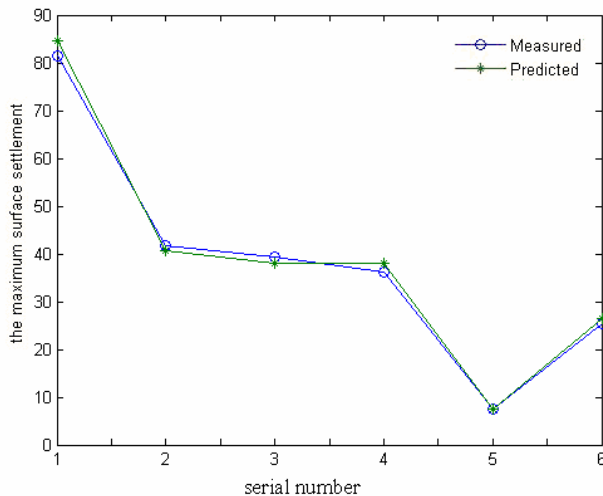
**Fig. 2.** Performance of neural networks on trained (predicted) and measured data

Table 2. Momentum coefficient is 0.9, steps to take 0.1, after e-learning  $2.2665 \times 10^5$  times, reach the target error of  $10^{-5}$ . The results from testing of validation set are plotted together in Fig. 1. As can be seen, the ANN predictions fit the data in the testing samples.

According to Table 3 and Figure 2, the maximum error of the predicted values and measured values of surface maximum settlement from three testing samples is less than 7%, which shows that the model's predicted performance is satisfactory within the permitted scope of the project.

### 3.4 Engineering Example

The Tunneling Project between Pudong South Road Station and Nanpu Bridge Station is an important component of Shanghai LRT Line 2 Project as well as a major project in Shanghai. The tunnel starts from the end well west to Pudong South Road Station to the end well east to Nanpu Bridge Station, with the full length of up line 1997.148 m and that of down line 1981.96 m, external diameter 6.2m and inner diameter 5.5 m. The shield external diameter is 6.34 m, the width of tunnel lining ring is 1m, the shield advancing speed is 2.0m, the total thrust of jack is 14m, the shield tail's grouting pressure is 0.3m, the grout is retarding grout whose intensity in 1d is greater than or equal to the intensity of the surrounding reinforced soil. The shield tail's grouting volume is  $2.0\text{m}^3/\text{m}$ . The tunnel settlement is expected at the in 1480 EMS-chip cross-section. The depth of the tunnel central axis at 1480 EMS-chip is 22m, overburden thickness 19m, mainly silty clay.

The surface maximum settlement of the 1480 ring predicted by the trained ANN model to is  $S_{\max}$  is 16.95mm, compared with the measured maximum surface settlement  $S_{\max}$  is 16.20 mm, only has a error of 4.63%.The predictive model performed well and confirmed that artificial neural networks can be successfully used for predicting surface settlements in the project.

## 4 Conclusions

The ANN system used in this paper demonstrated very satisfactory results in predicting the surface maximum settlement for the case study in question with confidence. The resulting remarks can be drawn hereinafter:

(1) ANN can decrease the errors to permitted targets ,in limited trained times ,by choosing proper self-adapted learning speed and momentum factors.

(2) Through continuous learning, ANN can use a simpler model to accurately predict the surface maximum settlement caused by shield tunnel excavation. This can form a stark contrast with the complex calculations in mechanical system. ANN has a strong anti-interference ability; individual forecasts will not affect the calculation results.

(3) By inputting the factors affecting the surface maximum settlement, the already trained ANN model can then accurately identify the surface maximum subsidence caused by shield tunnel excavation. The more training samples, the closer output gets to the fact.

(4) By analyzing and predicting the surface maximum settlement caused by shield tunnel in Shanghai with the method proposed in this paper, the results are consistent with the engineering practice. This further verifies the correctness of the model. It shows that BP ANN applying to the surface maximum settlement caused by shield tunnel has a broad application prospects.

## References

1. Peck, R.B.: Deep excavation sand tunneling in soft ground. In: Proc. 7th Int. Conference on Soil Mechanical and Foundation Engineering, Mexico City, pp. 225–290 (1969)
2. Fang, Y.S., Lin, S.J., Lin, J.S.: Time and settlement in EPB shield tunneling. *Tunnels & Tunneling* 134(11), 27–28 (1993)
3. Tang, Y., Ye, W., Zhang, Q.-h.: Analysis and research on ground settlement by the shield tunneling in Shanghai (three). *Underground Space* 15(4), 250–258 (1995)
4. Zhang, Y., Yin, Z., Xu, Y.-f.: Ground surface deformation by shield tunneling. *Journal of Rock Mechanics and Engineering* 21(3), 388–392 (2002) (in Chinese)
5. Kasper, T., Meschke, G.: On the influence of face pressure, grouting pressure and TBM design in soft ground tunneling. *Tunnelling and Underground Space Technology* (21), 160–171 (2006)
6. Karakus, M., Fowell, R.J.: Effects of different tunnel face advance excavation on the settlement by FEM. *Tunnelling and Underground Space Technology* (18), 513–523 (2003)
7. Liu, Z., Wang, M., Dong, X.: Analysis of ground surface settlement by construction with the shield tunneling method. *Journal of Rock Mechanics and Engineering* 22(8), 1297–1301 (2003) (In Chinese)
8. Huayang, D., Juecheng, Z., Youjian, H.: Testing study on surface displacement of mountainous region with similar material. *Journal of Rock Mechanics and Engineering* 19(4), 501–504 (2000) (in Chinese)
9. Guoxiang, Z.: Elasto-plastic and stochastic medium method and application in analysis of ground displacement and stress change due to tunnel construction. *Journal of Rock Mechanics and Engineering* 22(4), 596–600 (2003) (in Chinese)
10. Yi, H.-W., Sun, J.: Mechanism analysis of disturbance caused by shield tunneling on soft clays. *Journal of Tongji University* (3), 277–281 (2000) (in Chinese)
11. Sun, J., Yuan, J.-r.: Stratum movement and disturbance during the shield tunneling and their intelligent nerve net prediction. *Journal of Geotechnical Engineering* 23(5), 261–267 (2001)
12. Suwansawat, S., Einstein, H.H.: Artificial neural networks for predicting the maximum surface settlement caused by EPB shield tunneling. *Tunnelling and Underground Space Technology* (21), 133–150 (2006)
13. Santos Jr., O.J., Celestino, T.B.: Artificial neural networks analysis of São Paulo subway tunnel settlement data. *Tunnelling and Underground Space Technology* (23), 481–491 (2008)

# Applying Grey-Fuzzy-Fuzzy Rule to Machine Vision Based Unmanned Automatic Vehicles

Chin-Wen Chuang and Jau-Rong Li

<sup>1</sup> I-Shou University, Department of Electrical Engineering,  
1, Section 1, Hsueh-Cheng Rd., Ta-Hsu Hsiang, Kaohsiung County, Taiwan  
cwchuang@isu.edu.tw

<sup>2</sup> National Kaohsiung Normal University, Department of Business Management  
jrli@nknu.edu.tw

**Abstract.** The unmanned automatic vehicles (UAV) are important elements in industrial application, such as unmanned carriers in warehouse. With the great progress in the intelligent transportation systems, the UAVs are often applied to drive automatically. The mechanical vision replaces the human's eyes to be the feedback component. By the image processing, the necessary information can be figured out. However, due to the complicate calculation of image process, it will take a long calculation time from fetching image to making image information. Hence, a novel control rule that is the Grey-Fuzzy-Fuzzy controller is proposed in this paper for vision guided control of UAVs. The novel rule consists of the base-layer and upper-layer fuzzy controllers. The base-layer fuzzy controller is the classical fuzzy controller using the position error and the error difference to be the input variables. The Grey prediction is used to estimate the position error at the time of making image information. Then, the upper-layer fuzzy controller uses the two input variables, which are the estimated position error and the estimated error difference, to correct the base-layer fuzzy controller's output signal. Finally, an experimental UAV is developed to examine the potential of the proposed control scheme. The vision-based experiment of automatically driving of lane-following is carried out in this paper. The experimental results show the proposed controller will eliminate the swinging phenomenon and increase the accuracy while tracking. The experimental results also show the practical capability of the Grey-Fuzzy-Fuzzy controller.

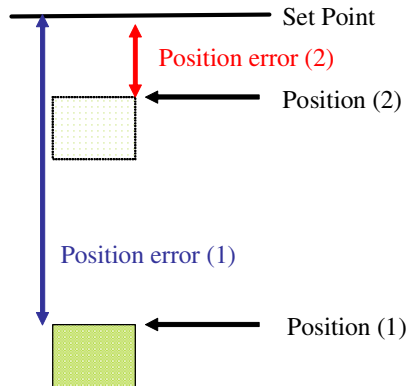
**Keywords:** Unmanned Automatic Vehicle; Grey Theory; Fuzzy Control; Grey-Fuzzy-Fuzzy Control.

## Introduction

The Unmanned Automatic Vehicle (UAV) is the important part in manufacture systems. The heavy productions are usually moved by the UAV. Traditionally, the UAV moves along the predefined road, usually marked by black taps. Hence, it is difficult to change the designed route. In recent decades, the machine vision is great developed. In robotics and automation, the application of vision guided has also been widely used in many practical aspects. Such as Ref. [1], the CCD is installed in the

operation robotic arm bottom. The fine-tuning makes the accuracy operation finished. The CCD is also used in the cross road to monitor the road-traffic [2]. In Ref. [3-4], the mobile robotic uses the mechanical vision to complete the target tracking. In Ref. [5], the unmanned vehicle successfully went on urban-like road. In these relevant researches, the vision-based application is proven to be feasible and practical.

In this paper, a machine vision based UAV is developed to examine the capability of vision-servo. A web cam is installed in the notebook. This notebook placed upon the UAV is used to handle the calculation of relevant image process. Use the image processing strategy to distinguish the obvious target at the both side of the lane. Then, find the middle point of targets of both side and follow it to finish the unmanned driving. By vision guidance, the UAV can cooperate with the flexible manufacturing systems and alternate the route anytime. However, due to the complicate calculation of image process, it will take a long calculation time from fetching image to making image information. As shown in Fig. 1, the position 1 is the position at the time of fetching image. The real position after image process is position 2. Hence, the real position error is position error 2 not position error 1. In the application of vision-based following, this incorrect of position error will cause the automatic vehicle to produce the swinging phenomenon [3-4]. In order to eliminate the swinging phenomenon, most of researches reduce the vehicles' speed. The other method is to reduce the revising rate of angle of vehicles while target tracking. No matter what kind of mentioned above methods, it will reduce the practicability while in application.



**Fig. 1.** Diagram of Error of the Image

In order to improve the above mentioned drawback, a novel two-layer fuzzy control rule is proposed in this paper to guide the UAV based on vision servo feedback. The base-layer controller consists of the classical fuzzy controller. The upper-layer controller consists of the grey prediction based fuzzy controller. The two-layer fuzzy controller will guide the UAV based on vision feedback under the conditions of not to decrease driving speed and the revising rate of angle. The proposed Grey-Fuzzy-Fuzzy Controller can successfully reduce the swinging phenomena and achieve the smooth driving for vision-based unmanned vehicle. The Fuzzy-Fuzzy control and

Grey-Fuzzy control are studied for many years [3-4, 10-11]. In Refs. [3-4], the three values of the predicted position, position and velocity are used for the index values of the fuzzy controller in the same layer. The other research uses the Grey Model to derive the fuzzy controller [10]. In Ref [11], the gains of the linear functional controller, consisted of the predicted value, the error and the error derivation are auto-tuned by the fuzzy rules [11]. In this paper, the novel method to combine the Grey theory and the fuzzy rule is proposed to apply to the image based unmanned auto-driving. The base-layer fuzzy controller is designed based on the classical fuzzy rule [3-4] [6-8], whose input variables are the position error and the error difference. In order to correct the error caused by the time spent by image processing, the upper-layer fuzzy controller is used to correct the base-layer fuzzy controller. The upper-layer consists of the two input variables, which are the estimated error and the estimated error difference. The Grey theory is used to predict the position error at the time of finishing image process. The Grey theory is proposed by Dr. Deng in 1982 [9]. Most of other predictive rules need to collect the long time-series data to make up the mathematic model. It will take lots of usage of memory and calculation time. Different from these rules, the Grey prediction needs only four time-series data to estimate the next position [3-4] [9-13]. The Grey theory is belonged to the piecewise estimation. Hence, this rule is suitable for on-line and simultaneous estimation work. The novel control rule will guide the AGV to drive on the lanes based on vision feedback without swinging phenomenon.

Finally, an experimental UAV is developed to examine the proposed control rule in this paper. The three control methods of the PID, the Fuzzy and the Grey-Fuzzy-Fuzzy are compared in these experiments. The experimental results show the good performance of vision-based automatically driving. The results also show the practical capability of the proposed rule in the intelligent transportation systems.

## Image Processing

The subject of the image processing is to figure out the necessary information from the grabbed image. In order to along the lane, the driving direction has to be decided firstly. Fortunately, the objects of the road and both sides have obvious or easy differences to come to conduct to distinguish, such as the asphalt road and meadow. Hence, the first step of image processing is to isolate the obviously distinguishing targets in the color image. The simplest method is to specify the threshold interval for each of the three color components. Then, color threshold converts a color image to a binary image. The pixels of the distinguishing targets are set to 1, the others are set to 0. According to the driving speed of vehicle, the search paths are specified to be two horizontal lines drawn across the image. The edge detection process will look for edges along the search paths among the binary image. Because of the width, there are two points found in the lane mark. The next step is to find the center pixel of the inner two edge point on the each search path. According the above mentioned method, the experimental result of image processing is shown in Fig. 2.



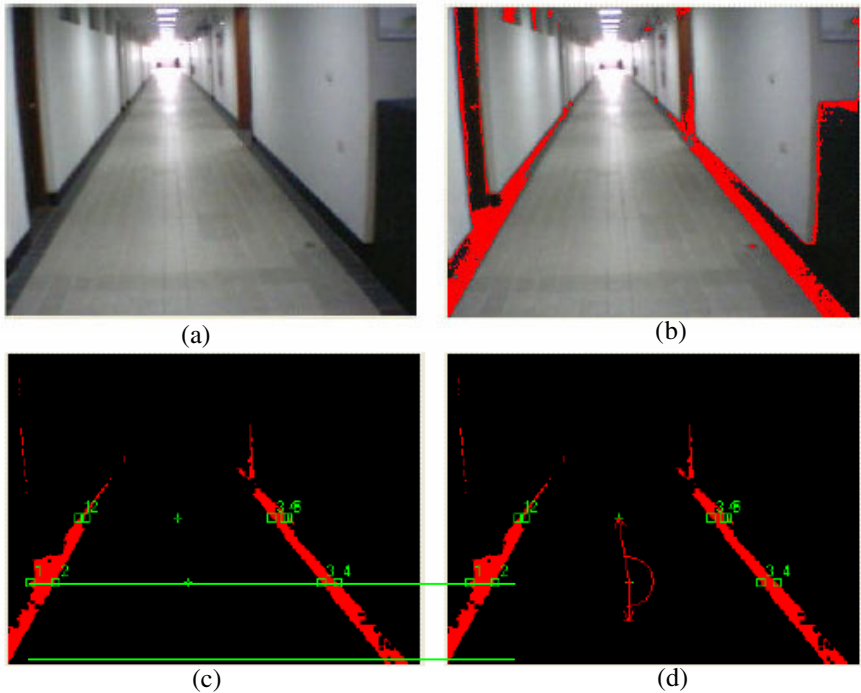


Fig. 2. Experimental Result of Image Processing

## Grey-Fuzzy-Fuzzy Control

### Fuzzy Control

The proposed controller is based on the classical fuzzy rule. Hence, the design procedures of the fuzzy rule are Fuzzification, Fuzzy Inference and Defuzzification. In order to fuzzify the input variables, the membership functions should be defined firstly. To simplify the relevant calculation, the triangular and trapezoidal membership functions are used in this paper. The range of the angle error from  $-75^\circ$  to  $+75^\circ$  is equally divided into five triangular membership functions and two trapezoidal membership functions. The range of the angle error difference ( $-50^\circ \sim +50^\circ$ ) and the output voltage ( $-10V \sim +10V$ ) of revising angle are divided by the same rules.

The Mamdani typed fuzzy inference is used in this paper, that is the MIN-MAX method. Use the linguistic description, that is the rule of If-Then, to decide the output relationship. The eq. (1) is one example of those descriptions.

$$\begin{aligned}
 &\text{IF } e(k) \text{ is negative but medium} \\
 &\quad \text{AND } [e(k) - e(k-1)] \text{ is about zero} \\
 &\text{THEN } u_i(k) \text{ should be positive but medium}
 \end{aligned} \tag{1}$$

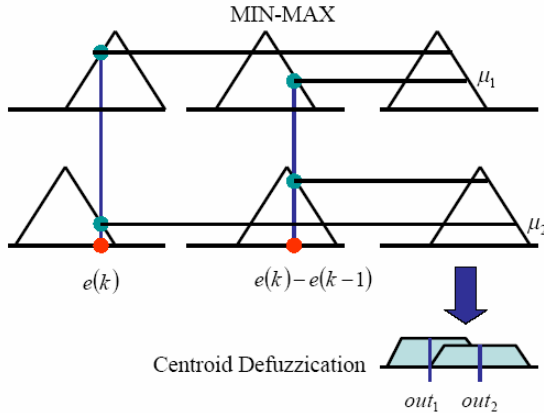


Fig. 3. Fuzzy Control Rule

Finally, by the Centroid method, the distinct output value can be obtained by eq. (2). The diagram of fuzzy rule is shown in Fig. 3.

$$u(k) = \frac{\sum \{out_i [1 - (1 - \mu_i)^2]\}}{[1 - (1 - \mu_i)^2]} \tag{2}$$

**Grey Prediction**

The significant advantage of the Grey Prediction theory is only used four time-series data in estimation process. Hence, the error and the data of three sampling time forward now consist of the data matrix, that is

$$S = [e(k-3), e(k-2), e(k-1), e(k)]$$

In order not to meet the singular point, a positive constant is added to the data matrix, that is

$$S^{(0)} = [e^{(0)}(k-3), e^{(0)}(k-2), e^{(0)}(k-1), e^{(0)}(k)]$$

where

$$e^{(0)}(i) = e(i) + \text{bias}$$

The Accumulated Generating Operation (AGO) is defined in Ref. [9] to approximate the integral operation by accumulated summation. The Inverse Accumulated Generating Operation (IAGO) is the inverse operation of the integral approximation. According to the AGO and IAGO, the time-series data can be described as

$$e^{(1)}(i) = \text{AGO} \bullet e^{(0)}(i)$$

where

$$e^{(1)}(k-3+i) = \sum_{j=3}^i e^{(0)}(k-3+j) \quad \text{for } i=0,1,2,3$$

and

$$e^{(0)}(i) = \text{IAGO} \bullet e^{(1)}(i)$$

$$e^{(0)}(k-i) = e^{(1)}(k-i) - e^{(1)}(k-i-1) \quad \text{for } i \leq 2$$

$$e^{(0)}(k-3) = e^{(1)}(k-3) \quad \text{for } i = 3$$

In Ref. [9], the Grey Model which is the first order and has one variable is defined as eq. (1).

$$e^{(1)}(k-i) + a m^{(1)}(k-i) = b \tag{1}$$

where

$$m^{(1)}(k-i+1) = \frac{1}{2} [e^{(1)}(k-i+1) + e^{(1)}(k-i)] \quad \text{for } i=3,2,1$$

By the Least-Square method, the parameters of the Grey Model can be figured out, like eq. (2).

$$\begin{bmatrix} a \\ b \end{bmatrix} = \left( \begin{bmatrix} -m^{(1)}(k-2) & 1 \\ -m^{(1)}(k-1) & 1 \\ -m^{(1)}(k) & 1 \end{bmatrix}^T \begin{bmatrix} -m^{(1)}(k-2) & 1 \\ -m^{(1)}(k-1) & 1 \\ -m^{(1)}(k) & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} -m^{(1)}(k-2) & 1 \\ -m^{(1)}(k-1) & 1 \\ -m^{(1)}(k) & 1 \end{bmatrix}^T \begin{bmatrix} E^{(0)}(k-2) \\ E^{(0)}(k-1) \\ E^{(0)}(k) \end{bmatrix} \tag{2}$$

Once obtaining the parameters of the Grey Model, the next status can be estimated like eq. (3).

$$\tilde{e}^{(0)}(k+1) = \frac{b - a \times e^{(0)}(k)}{1 + 0.5 \times a} \tag{3}$$

and

$$\tilde{e}(k+1) = \tilde{e}^{(0)}(k+1) - \text{Bias}$$

**The Grey-Fuzzy-Fuzzy Control**

The novel control scheme is developed based on the above mentioned control rules. The error  $e(k)$  and the error difference  $[e(k) - e(k-1)]$  compose the base-layer fuzzy controller. The base-fuzzy controller plays the main role of generating basic guidance signal  $u^0(k)$ . Due to the error caused by the time from fetching image to making image information, the Grey prediction is used to compensate the possible error. The estimated error  $\tilde{e}(k+1)$  means the error at the time of making image information.

Then, by the estimated error  $\tilde{e}(k+1)$  and the estimated error difference  $[\tilde{e}(k+1)-e(k)]$ , the upper-layer fuzzy controller figures out the correcting signal  $\Delta u(k)$  to correct the base-layer fuzzy controller. The detailed diagram of the Grey-Fuzzy-Fuzzy control is shown in Fig. 4. The relevant signals such as feedback and output are restricted by the hardware limitation. While applying the signal into the system within the range, the closed loop system can be still guaranteed to be stable according to the experimental results. Therefore, the close loop system can be concluded to be stable by the theory of Uniformly Ultimately Bounded [14-15].

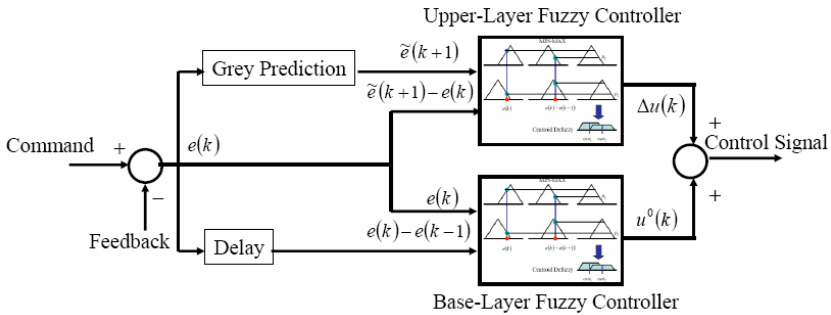


Fig. 4. Diagram of Grey-Fuzzy-Fuzzy Control

### Experimental UAV

In order to examine the potential of the proposed control scheme, an experimental UAV is developed in this paper. The hardware diagram is shown in Fig. 5.

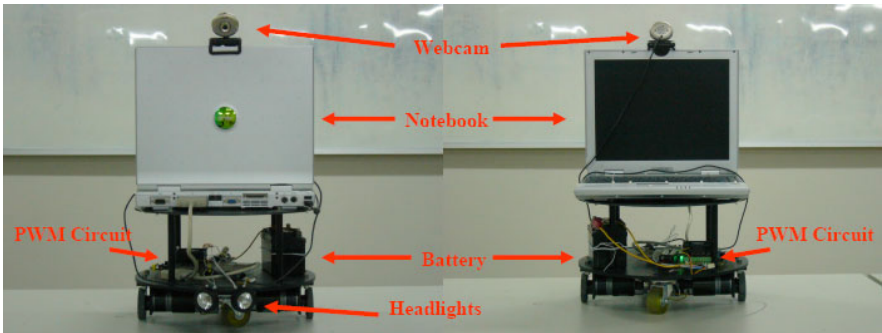


Fig. 5. Experimental UAV

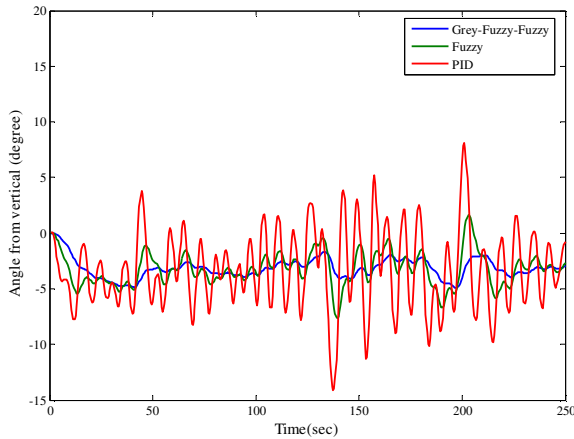
The notebook of the Pentium grade deals with the hardware input/output control and the necessary software calculation of the image processing and the proposed controller. The human machine interface is programmed by the software of

LabVIEW<sup>1</sup> and Matlab/Simulink<sup>2</sup>. A web cam is connected with the notebook by the USB connector, which is used to grab image of front direction of vehicle. Through the printer port, the 8-bits pulse width modulator prototype board is used to control the both DC motors drive the both-side wheels. Two headlights are installed in the front of the vehicle. They will be turn-on to increase the color contrast when the driving direction detection failed.

## Experimental Results

The experiment is carried out in the corridor of the teaching building of I-Shou University in Taiwan. The moving ahead speed of the UAV is maintained at the maximum speed that is 2.5km/hr. By mechanical vision-based distinguishing the baseboard of both side-walls, the UAV moves along the center of the corridor.

Three controllers of Grey-Fuzzy-Fuzzy, Fuzzy and PID are compared under different environment. The parameters of the PID controller are:  $K_p = 3$ ,  $K_I = 0.1$  and  $K_D = 0.3$ . When the light is bright and the AGV walks along the clean corridor, the experimental results are shown in Fig. 6. Obviously, the Grey-Fuzzy-Fuzzy controller can offer the more comfortable auto-driving. The Fuzzy controller also has the perfect performance. However, the PID controller will shows the great swinging driving phenomenon.

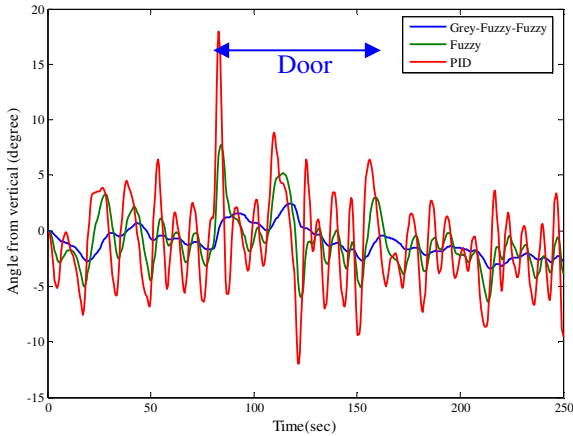


**Fig. 6.** Experimental Result with clean corridor and bright environment

When the UAV walks along the corridor with a deep color door, the image process will increase the calculation time for turning-on the headlights and re-sampling. The experimental results are shown in Fig. 7. The PID and Fuzzy controllers will have the bigger jumping points in driving path. The proposed controller still has smooth driving path.

<sup>1</sup> The LabVIEW is the trademark of NI CO.

<sup>2</sup> Matlab and Simulink are the trademark of Mathworks CO.



**Fig. 7.** Experimental Result with deep color door in corridor and bright environment

## Conclusion

The Grey-based two-layer fuzzy controller is applied to vision guided UAV. The detailed design procedures are described in this paper. Due to the long time from fetching image to make image information, applying the classical controller, such as the PID controller, will have the swinging phenomenon. Hence, the Grey prediction is used to estimate the position at the time finishing making image information. Combining the Grey prediction with the fuzzy control rule, the novel two-layer fuzzy controller is proposed in this paper. The experimental results show the significant performance in eliminating the swinging phenomenon. More smoothly following lane driven can be observed in the experimental results compared with the PID and Fuzzy controllers.

## Acknowledgement

This paper is partial financial supported by the National Science Council of the project number NSC98-2220-E-214-001.

## References

1. Ralis, S.J., Vikramaditya, B., Nelson, B.J.: Micropositioning of a Weakly Calibrated Microassembly System Using Coarse-to-Fine Visual Servoing Strategies. *IEEE Trans. on Electronics Packing Manufacturing*, 123–131 (2000)
2. Setchell, C., Dagless, E.L.: Vision-Based Road-Traffic Monitoring Sensor. *IEE Proc.-Vis. Image Signal Process*, 78–84 (2001)
3. Luo, R.C., Chen, T.M.: Autonomous Mobile Target Tracking System Based on Grey-Fuzzy Control Algorithm. *IEEE Trans. on Industrial Electronics*, 920–931 (2000)

4. Luo, R.C., Chen, T.M., Su, K.L.: Target Tracking Using Hierarchical Grey-Fuzzy Motion Decision-Making Method. *IEEE Trans. on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 179–186 (2001)
5. Sotelo, M.A., Rodriguez, F.J., Magdalena, L.: VIRTUOUS: Vision-Based Road Transportation for Unmanned Operation on Urban-Like Scenarios. *IEEE Trans. on Intelligent Transportation Systems*, 69–83 (2004)
6. Zadeh, L.A.: Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. on Systems, Man, and Cybernetics*, 28–44 (1973)
7. Lee, C.C.: Fuzzy logic in control systems: fuzzy logic controller – Part I. *IEEE Trans. on Systems, Man, and Cybernetics*, 401–418 (1990)
8. Lee, C.C.: Fuzzy logic in control systems: fuzzy logic controller – Part II. *IEEE Trans. on Systems, Man, and Cybernetics*, 419–435 (1990)
9. Deng, J.L.: Control problems of grey systems. *Systems & Control Letters*, 288–294 (1982)
10. Feng, H.M., Wong, C.C.: An on-line rule tuning grey prediction fuzzy control system design. *Proceeding of IEEE*, 1316–1321 (2002)
11. Huang, Y.P., Huang, C.C.: The integration and application of fuzzy and grey modeling method. *Fuzzy Sets and Systems*, 107–119 (1996)
12. Chuang, C.W., Chou, C.H.: Motion Control of X-Y Table by Sliding Mode Control with Position Compensator in Feedback Loop. In: *The Sixth International Conference on Automation Technology*, pp. 677–682 (2000)
13. Huang, S.J., Huang, C.L.: Control of an Inverted Pendulum Using Grey Prediction Model. *IEEE Trans. on Industry Applications*, 452–458 (2000)
14. Lee, Y., Zak, S.H.: Uniformly Ultimately Bounded Fuzzy Adaptive Tracking Controllers for Uncertain Systems. *IEEE Transactions on Fuzzy Systems*, 797–811 (2004)
15. Zhang, X., Fan, Y.: An Algorithm of Uniform Ultimate Boundedness for Switched Linear Systems. In: *IEEE SMC 2003 Conference Proceedings*, pp. 166–170 (2003)

# AWARE: Autonomous Wireless Agent Robotic Exchange

Cory Q. Nguyen<sup>1</sup>, Benny Leong<sup>1</sup>, Eric T. Matson<sup>1</sup>,  
Anthony Smith<sup>1</sup>, and Juan P. Wachs<sup>2</sup>

<sup>1</sup> Computer and Information Technology

<sup>2</sup> Industrial Engineering

Purdue University, West Lafayette, Indiana, USA

{cqnguyen,bleong,ematson,ahsmith,jpwachs}@purdue.edu

<http://m2m.tech.purdue.edu>

**Abstract.** The Autonomous Wireless Robotic Agent Exchange (*AWARE*) system revolves around autonomous self-organizing wireless networks that provide end-to-end communication via a wireless medium. It provides an immediate communication solution to disaster stricken regions and to inaccessible geographical areas where wireless communication is the only viable option, but most current systems suffer due to their static nature. *AWARE* has many real-world applications related to search and rescue, disaster recovery, military and commercial industries.

## 1 Introduction

In one part of the world, we comfortably go through our work day with seemingly unlimited access to information and communications. We take for granted the conveniences of the wireless technologies that provide us these luxuries; it has become an expectation, a basic expected commodity. In another part of the world, a strong earthquake strikes Port-au-Prince, Haiti, roughly 230,000 people are left dead, and another 1,000,000 left homeless [1]. Chaos and petty crimes plague the once vibrant city, as cries of anguish and bellowing calls for help are heard abound. To compound the immediate problems, all communication infrastructures are lost. Gone are the conveniences of cell phones and wireless Internet connection, people are left stuck and stranded, waiting for help.

If only people, in a situation similar to this, had some way to contact help. If only they had the ability to reach out to emergency personnel or neighboring cities alerting them to those needing medical attention. Better communication can facilitate the relief effort among the rescue teams and significantly reduced the death toll [2]. Imagine, amongst all the chaos, destruction, and devastation, the people had some way of establishing communication, either through the use of VoIP, e-mail, or other messaging services using a wireless network backbone that was instantly deployed after the earthquake.

The research presented in the paper offers the conceptualization of the initial stages and results that can ultimately lead to the envisioned solution, an autonomous robotic wireless mobile self-configuring network that can be deployed



almost anywhere in the world, to any situation. This technology will provide communication links and Internet access to parts of the world where wireless did not or could not exist before. This technology can be adopted in the areas of public safety, search and rescue, and military or government operations. This paper describes our initial work and the early prototype of our *AWARE* research.

## 1.1 Background

The concept of controlling a robot wirelessly from a remote site is not a new idea. Similar technologies have been implemented in drones used in military applications [3], bomb remediation robots [4] and space agencies [6].

The envisioned concept of autonomous wireless robotics, presented in this paper, is entirely different from previous examples of wirelessly controlled robots with two important distinctions. In our research [10], the element of autonomy is required. Secondly, instead of the use of traditional wireless ad-hoc network for robotic communication, wireless mesh technology is utilized instead.

The envisioned concept of *AWARE* system, besides being physically deployed by humans, the *AWARE* robots possess full control over their course of actions. While being fully autonomous, the robots are capable of self organizing and self healing the wireless network while collaborating accordingly amongst each other based on their environment in order to establish a mobile wireless backbone.

The wireless backbone in this research is constructed using wireless mesh technology instead of the traditional wireless ad-hoc networking [7]. A wireless ad-hoc network, although extremely convenient, possesses a great limitation especially in large scale applications. The network throughput of a wireless ad-hoc network suffers a great performance drop as more users are connected to the network. Data has shown network throughput drops exponentially as more users connect to the network [8]. On the other hand, in a mesh network, every single node is able to connect and communicate with each other; thus, creating a network redundancy. So, by design, the aforementioned limitation does not apply to your research.

In another study, the One Laptop Per Child (OLPC) uses mesh networking to connect laptops to each other with several limitations. The mesh network only allows similar laptops to connect to the network and nothing else. Additionally, it lacks mobility as the network solely depends upon the people to establish a connection [9]. The mobile wireless robots, in this paper, establishes an autonomous wireless mesh network that will allow any device or user to connect and access the network, eliminating device limitations and the dependency on people.

This research combines the benefits of mesh networking, like the OLPC project, and the advantages of autonomous self-organization to foster a new concept of an autonomous mobile wireless network that will be able to cover and span geographical distances and can be applied in most task environments.

The objective of the autonomous wireless mobile robots is to establish and maintain a mobile wireless backbone that is both scalable and adaptable to its ever-changing environment. The application for such a technology like mobile wireless backbone is limitless. To list a few examples of applicability, we turn to scenario-based situations.

First of all, in an event of an earthquake disaster like Port-au-Prince, Haiti, a mobile wireless backbone is useful because it will be able to navigate the heavily damaged terrain and provide a wireless network that would be able to cover many critical rescue sites. With the wireless backbone established, injured civilians can establish contact with emergency personnel using all forms of network-based wireless communication, such as VoIP, e-mail or texting. Additionally, rescuers and first responders can communicate to coordinate evacuation, search and rescue missions more effectively. Search and rescue robots can also be deployed and utilize the established wireless mobile mesh backbone to communicate and receive instructions from central command [11].

In the event of a mine-related disaster, a wirelessly controlled robot is deemed useless in search and rescue efforts due to the inability of a wireless signal to penetrate thick, solid rock walls. However, with autonomous wireless mobile robots, a wireless backbone can be established and extended into the mine in order to enable the application of wirelessly controlled robots and communication to lost miners.

Besides disaster relief, a mobile wireless backbone can also be valuable to researchers. In particular, scientists who need to collect environmental data and geological data under extreme conditions [5]. For instance, a scientist in Antarctica has to travel an hour or two from their home base in order to collect ice samples or temperature readings. With a mobile, reconfigurable wireless backbone, specialized robots can be deployed and controlled from the home base. Then, using the same wireless backbone, data collected by the specialized robots can be transmitted back to the scientist back in the home base.

Maintaining a communication link from a robot with its remote control station has been a problem for both military and law enforcement operations. There has been research that has attempted to address this issue; however, they present limitations with the wireless technology and methodology used, especially when employing Ad-hoc technology as a means to establish end-to-end communication via wireless [12]. The presented autonomous mobile wireless mesh network established by the autonomous wireless mobile robots not only addresses the limitation that previous research attempt to solve but also provides a scalable, adaptable, and redundant wireless backbone capable of supporting various application under different situation or environment. The aforementioned scenarios are the few of many where this mobile wireless backbone can be applicable and useful to the society.

In section 2 we show the concept of the *AWARE* system and its capabilities. In section 3, the realization and evaluation are discussed. In section 4, conclusions to this research are given and we discuss some future plans to extend this work.

## 2 Concept

The *AWARE* system includes multi-agent communication and coordination. The goal of the agents is to collaborate with one and another to form the desired optimal wireless coverage in the form of a mesh environment or a point-to-point environment. The completion of the *AWARE* prototype allows the robots to form a fully functionally self-organized mobile wireless network.

The agent of each unit is designed to collaborate and work together. There will be at times when agents establish roles to get the job done. For example, a designate leader is determined and all followers will perform tasks assigned by the leader. At other times, the agents are individual autonomous units working with other individual autonomous agents, collaborating and coordinating with one another to get the job done.

### 2.1 Intelligent Wireless Mesh Robotics

The purpose of *AWARE* is to provide quick and immediate wireless coverage to a geographical area that would require communication or Internet access where it currently does not exist.

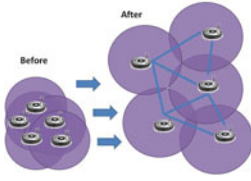
The agents of the wireless mesh robotics environment work with one another to optimize the wireless coverage transforming from a small cluster to an expanded formation covering a large geographical area. It will also be able to automatically self-heal and reconfigure the network accordingly, especially in cases where a robot unit malfunctions or is destroyed.

To do so, the agents rely on GPS and a digital compass module along with an expanded algorithm to be able to coordinate with one another to determine which agents will remain stationary and which agents are allowed to move about. They will be able to tell each other where to go, how far to go, and even how fast to drive. The agents will apply their autonomous ability to navigate their environment, avoiding obstacles. As shown in Fig. 11, the agents work together to expand their coverage area while still maintaining communication with each other. They will continuously change modes to accommodate and situate themselves in the right place to provide the necessary wireless coverage.

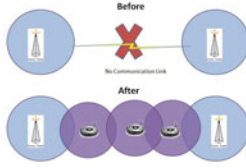
Please note that *AWARE* would rely on GPS to map its's position for use of outdoor scenarios, however, when *AWARE* is deployed in indoor environments, it will rely on a variety of sensors and localized communication technologies to be able to navigate its way about obstacles to fulfill its objective of providing optimized wireless coverage.

### 2.2 Wireless Distribution System (WDS) Robotics

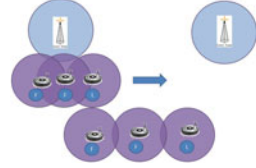
The *WDS* environment requires the agents to work together to expand in a linear formation to complete an end-to-end communication link. At times, situations and scenarios calls for a complete end-to-end link versus a mesh like coverage area. In an end-to-end link scenario, fewer agents are required to accomplish the



**Fig. 1.** Concept



**Fig. 2.** Linear Formation



**Fig. 3.** Accordion Effect

job. Fig. 2 illustrates the use of a *WDS* configuration versus a wireless network is to complete the link from point A to point B.

The algorithm to accomplish this task utilizes a follow the leader concept. The agents all start out in a straight line, with the leader at the very front. The leader will start driving towards destination B. As the leader continues to drive, the last agent of the convoy will begin to track its RSSI with the wireless tower at destination A.

When it reaches a certain threshold, the agent will stop and notify the convoy that it has stopped. The convoy then calculates how many agents are left and the new agent at the back of the convoy is determined. The end node agent will repeat the algorithm of its predecessor and stop when a certain RSSI threshold is reached. This process continues until the convoy can stretch the link no more, or the link has been successfully completed.

Fig. 3 shows the accordion effect, where the leader continues to drive and the followers are right behind but stopping only when they cannot continue further or else they would break the wireless communication link.

The robots are able to follow the leader using infrared technology. The robot will transmit an infrared signal from the back; the following robot will be able to pick up on the infrared signal and will continue to drive as long as it sees the infrared signal.

As it continues to follow, it will constantly read the RSSI to make sure it has not reached the threshold, as soon as it does, it will stop and notify the rest of the convoy. As described earlier, the convoy will recalibrate and continue to proceed.

### 3 Realization and Evaluation

To better understand the *AWARE* system, the anatomy and use of agents is shown in the this section.

#### 3.1 Anatomy

The wireless autonomous mobile robot is composed of three elements: a wireless interface, a processing unit, and a mobile transport. The concept is shown in Fig. 4 and the robotic instance is shown in Fig. 5.



Fig. 4. Concept



Fig. 5. Robotic Instance

**Wireless Interface.** The wireless interface allows the processing unit to collect the necessary wireless data for decision making. The wireless data in this case is the received signal strength indicator (RSSI), which determines the signal strength of a wireless connection. The RSSI values for the Proxim AP-4000 range between 0 and 100, with higher RSSI values indicating better signal strength. At an RSSI of 50 or more, the connection data rate is between 48 Mbps and 54 Mbps, which is considered the acceptable range, in this research.

**Processing Unit.** Based on the collected RSSI, the processing unit, which is the Dell Mini 10 net book, makes a variety of decisions based on its designed algorithm. For example, when the processing unit receives a RSSI of 50 or higher, it would issue a stop command to the mobile transport. Otherwise, it would continue to issue a drive command to the mobile robot.

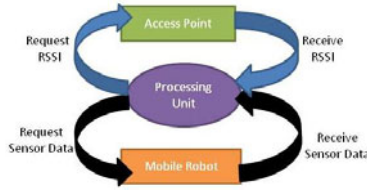
**Mobile Interface.** In order for the robot to know where to move and what to do if it is caught in a sticky situation, the processing unit communicates specific instructions to the mobile transport, giving it simple instructions: to turn right, left, drive straight, or back up. The instructions sent are dependent upon both the RSSI that is collected from the wireless interface and the sensor data received from the mobile transport. The RSSI and sensor data are precepts that allow the robotic agent to make the appropriate decision. The processing unit communicates to the mobile transport via a Bluetooth connection.

The mobile transport used for this research is the iRobot Create, which is equipped with front and side bumper sensors as well as actuator motors that allow it to drive and turn. Additionally, it has cliff sensors to help the robot determine whether it is going to drive off a cliff.

**Internal Communication.** As stated earlier, the various components of the wireless autonomous mobile robot communicate with the processing unit via serial port using USB-to-Serial converters and Bluetooth adapters. Fig. 6 illustrates the internal communication for the *AWARE* robotic instance.

The communication process is continuous as the processing unit is constantly transmitting to and receiving from both the access point and mobile transport.

The communication between the processing unit and the wireless interface is solely for the purpose of collecting RSSI data. The processing unit would send a



**Fig. 6.** Internal Communication Flows

command to request for the RSSI of the access point, which would then respond accordingly. USB-to-Serial converter is used as a connection medium between the wireless interface and the processing unit.

Thereafter, based on the collected RSSI, the processing unit would make a decision based on the algorithm for establishing wireless communication. Once the appropriate action is determined, the processing unit sends commands to the transport vehicle to either move in a specific direction or stop.

### 3.2 Autonomous Agent Approach

To evaluate the best action, both goal-based and utility-based agent constructs were considered. The goal of the goal-based agent was to achieve a RSSI of 50 or more. Unless the goal of the agent is achieved, the agent would continue to move forward. Any changes in the RSSI value would cause the agent to re-evaluate its course of action. Table 1 below summarizes the information on this model.

**Table 1.** Goal-based Agent

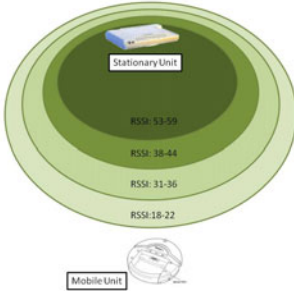
Prerequisite	RSSI target: 50 or higher
Objective	Locate an area with AP coverage
Course of Action	<ol style="list-style-type: none"> <li>1. Continuously moving forward until a RSSI value is found</li> <li>2. Stop moving as soon as a RSSI value is found</li> <li>3. Continuously monitoring RSSI value for changes</li> <li>4. Repeat Step 1 if RSSI value is lower than RSSI target</li> </ol>

The utility-based agent was designed to stop within an acceptable pre-defined RSSI range. Any changes in the RSSI value would cause the agent to re-evaluate its course of action. Table 2 in the following summarizes computation of the utility-based agent.

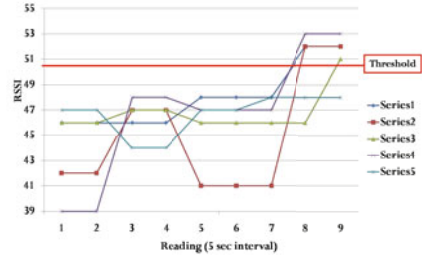
It was found the utility-based agent model works best for an autonomous wireless network due to a characteristic of the wireless frequency called multipath. Multipath causes wireless signals to be inconsistent even for stationary clients. The RSSI value would fluctuate on average a difference of 5 to 10. Hence, utility-based agent model gives leeway in its acceptance of what is considered a satisfactory RSSI. Fig. 7 shows how The autonomous mobile agent searches for and maintains the desired specified RSSI connection established with the located AP.

**Table 2.** Utility-based Agent

Prerequisite	RSSI range: 30-50
Objective	Locate an area with optimal AP coverage
Course of Action	<ol style="list-style-type: none"> <li>1. Continuously moving forward until a RSSI value within the RSSI range is found</li> <li>2. Stop moving as soon as a RSSI value within the RSSI range is found</li> <li>3. Continuously monitoring RSSI value for changes</li> <li>4. Repeat Step 1 if RSSI value is out of the RSSI range</li> </ol>



**Fig. 7.** RSSI Ranges



**Fig. 8.** Initial execution test results

Initial testing placed the autonomous mobile agent 10 feet away from the stationary AP. The agent was activated and begun to drive in a straight line until it received an RSSI of 50 or more. Once the agent received an RSSI of 50 or more, the agent would stop and continue to collect the RSSI from the AP. As soon as the RSSI goes below 50, the agent would continue to drive again. Based on the graph in Fig. 8, it is shown that the average RSSI reading between the 15th to 25th second time interval is around 45 for all 5 tests.

As the robot drives closer to the AP, the RSSI continues to fluctuate from a range of 41 to 48. However, the average RSSI from the data is around 47 RSSI. Finally after about the 35th second, a consistent RSSI reading of 50 is established. It is further proof that multipath creates an inconsistency in RSSI readings and is accommodated by using the utility-based agent.

### 3.3 End-To-End Communication

Once the agent was able to respond and perform its basic actions based on the readings of just one AP, the agent was extended with the ability to establish an end-to-end communication link with the use of two APs located at opposite ends of the room. Thereafter, the robot simply moved into a position where the APs could all see each other and be able to send traffic from one end of the network to the other end. The precondition is that the robot is placed at the center between the two APs, all it has to do is drive straight and detect that it has completed the link.

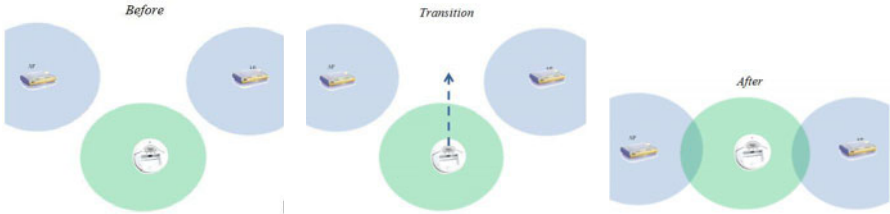


Fig. 9. Before

Fig. 10. Transition

Fig. 11. After

The importance of the agents' ability to detect multiple APs and be able to act accordingly is crucial to developing the fully mature autonomous wireless mesh network. By proving that an autonomous robot was able to successfully provide end-to-end communication, the accomplishment was an enforcement to the pragmatism of autonomous wireless mobile robots.

Fig. 9 shows AP on each end, currently unable to communicate with each other because they are out the range of each other. In Fig. 10, the agent will now move into the middle to complete the link and establish an end-to-end connection, Finally in Fig. 11, the agent drove in between the two APs to successfully complete the connection and provide end-to-end connectivity.

### 3.4 Optimization

The objective is to enable the agent to be fully autonomous and giving it the ability to adjust accordingly to its environment. Fig. 12 illustrates the basic optimization.

In order to prevent the agent from driving continuously into an obstacle, it was given the ability to read sensor data. It was able to detect when it bumped into a wall or an object and be able to turn in a different direction to continue driving. The direction it turned depended upon where it was bumped. For example, if it was bumped on right side, it would turn left 45 degrees. If it was bumped on left side, it would turn right 45 degrees. A common outcome is the signal strength not evenly distributed when the agent establish an end-to-end wireless communication. One end would end up with a higher RSSI compared to the

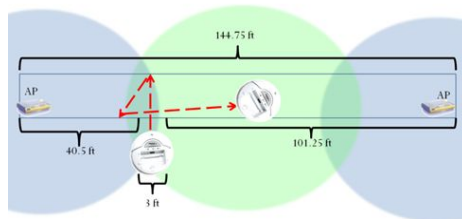


Fig. 12. Link optimization



**Table 3.** Optimization Data Example

2010-03-27 10:37:05	No RSSI.
2010-03-27 10:37:18	RSSI 1: 6.
2010-03-27 10:37:37	No RSSI.
2010-03-27 10:37:45	RSSI 1: 15 RSSI 2: 17 found
2010-03-27 10:37:45	RSSI 1: 15 RSSI 2: 17 optimal

other end. This indicates one side of the connection could receive a data rate of 18 Mbps while the other could receive a data rate of 48 Mbps. This scenario is undesirable and we consider it to be non-optimal.

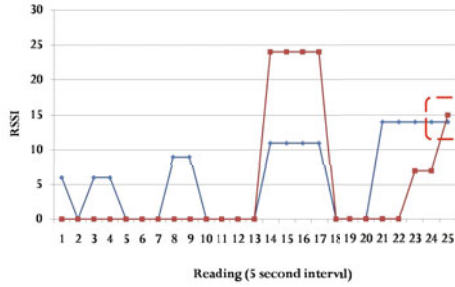
To remedy this problem, the agent was given the ability to optimize the wireless connection that it established with the APs it located. A link connection is considered optimal when the differences of the RSSI of the connected APs are within a defined acceptable range. From the example above, the optimization mode allowed the robot to have a 24Mbps and 36Mbps connection versus the 18Mbps and 48Mbps connection. This optimization process allows for a more stable connection.

Table 3 shows the data collected demonstrating the optimization mode of the robot. It began in a situation where it could not detect any RSSI from any of the APs. It started to drive straight to a point where it found the first AP with a very low RSSI. Notice that due to multipath, the signal from the first AP jumps around from an RSSI of about 6 to 9 to none. This shows that the link to the first AP was not very stable. Also at this point it has yet to find the second AP, it currently only shows the RSSI of just one AP.

As the robot moved about, it picked up both APs. At this point, it was in the discovery mode where it just found the two APs. It then entered optimal mode where it checked to see if the two RSSI were within the allowed range. The range set for this experiment run was with an RSSI difference of 5. The robot stopped as soon as it detected two RSSI values within the pre-defined range.

However, as it was listening, the signal dropped so it began driving around until it detected the two APs again. The signal drop off could be caused by many factors, such as interference from environmental noise or from a newly placed obstacle around the AP. In the case of this experiment, it was simply due to human movement, which caused a change in the environment for the robot. Even with the brief signal drop, it was able to find the two APs again and enter optimal range as shown in Fig. 13.

The blue line represents the RSSI activity between the agent and one AP, the red line represents the RSSI activity between the agent and of the second AP. During the beginning of the run, the agent only picked up one AP indicated by the blue line. The second AP was not detected until 13 seconds into the run of which it had an RSSI read of about 25. Now that the agent had located the two AP and established an end-to-end link, it began to optimize the link and moved about until both AP were within an RSSI of 5, indicated by the red circle on the graph. The completion of the optimization algorithm enables the agent



**Fig. 13.** Optimize after establishing end-to-end connection

to become adaptive to its environment and allows it to accomplish its goal of establishing an end-to-end connection upon detecting two APs.

## 4 Conclusions and Further Work

This research presents a novel combination of wireless mesh technology and robotics agents in creating a general-purpose autonomous wireless network with numerous applications. Even though results provided do not show a full implementation of the *AWARE* system, our initial results, and findings are, however, very promising. Our initial, prototypical work has provided insights on the advantages of wireless mesh technology compared to the ad-hoc connections used in many traditional robotic applications. Additionally, our work demonstrates the capability of a single robotic instance of the *AWARE* system. The successful demonstration of a single robotic instance lays the foundation and groundwork of which the *AWARE* system will be built upon, including the deployment of multiple robotic instances applying multi-agent cooperation and communication.

Our research is far from being done as there are many more areas and issues to be explored and enhanced. As presented in this paper, the current *AWARE* system relies on only one mobile robotic instance with cooperation of two stationary units. Future work for this research includes but are not limited to enhancing the coordination and cooperation of the system in a multi-agent environment, exploring positioning technologies such as infrared, agent orientation, and GPS, as well as extension to larger scale all-terrain robots.

The logical next steps will extend the current prototype to an outdoor robotic system using the current radio systems. This will allow a longer distance network to be formed. Then, the use of directional antennas will be employed to eventually create an *AWARE* system with a range of many miles over very complex terrain.

## References

1. Washington Post, Haiti raises earthquake toll to 230,000 (February 10, 2010), <http://www.washingtonpost.com>
2. Public health risk assessment and interventions. Haiti, Earthquake (January 2010), [http://whqlibdoc.who.int/hq/2010/WHO\\_HSE\\_GAR\\_DCE\\_2010.1\\_eng.pdf](http://whqlibdoc.who.int/hq/2010/WHO_HSE_GAR_DCE_2010.1_eng.pdf)

3. Predator RQ-1 / MQ-1 / MQ-9 Reaper - Unmanned Aerial Vehicle (UAV), USA, Airforce Technology,  
<http://www.airforce-technology.com/projects/predator/>
4. Jian-Jun, Z., Ru-Qing, Y., Wei-Jun, Z., Xin-Hua, W., Jun, Q.: Research on Semi-automatic Bomb Fetching for an EOD Robot. *International Journal of Advanced Robotic Systems* 4(2) (2007)
5. Nguyen, H.G., Pezeshkian, N., Gupta, A., Farrington, N.: Maintaining Communication Link for a Robot Operating in a Hazardous Environment. In: *ANS 10th International Conference On Robotics and Remote Systems for Hazardous Environments*, Gainesville, FL, March 28-31 (2004)
6. Jet Propulsion Laboratory, Mars Exploration Rovers, NASA (October 2004)
7. Timotheou, S., Loukas, G.: Autonomous Networked Robots for the Establishment of Wireless Communication in Uncertain Emergency Response Scenario. In: *Stochastic Agents: SAC 2009*, Honolulu, Hawaii, U.S.A, March 8-12 (2009)
8. Holland, G., Vaidya, N.: Analysis of TCP Performance over Mobile Ad Hoc Networks. *Wireless Networks* 8(2-3), 275–288 (2002)
9. Buchele, S.F.: Using OLPC Laptop Technology as a Computer Science Case Study Learning Tool. *The Journal of Computing Sciences in Colleges* 24(4), 130–136 (2009)
10. Smith, A., Matson, E.T.: Use of Antennas as Sensors to Discover Signals to Form Mobile Broadband Networks. In: *Proceeding of the 2009 IEEE Sensors Conference*, Christchurch, New Zealand, October 25-28 (2009)
11. Ko, A., Lau, H.Y.K.: Robot Assisted Emergency Search and Rescue System With a Wireless Sensor Network. *International Journal of Advanced Science and Technology* 3 (February 2009)
12. Pezeshkian, N., Nguyen, H.G., Burmeister, A.: Unmanned Ground Vehicle Radio Relay Deployment System For Non-line-of-sight Operations. In: *13th IASTED International Conference on Robotics and Applications*, Würzburg, Germany, August 29-31 (2007)

# Application of Wearable Miniature Non-invasive Sensory System in Human Locomotion Using Soft Computing Algorithm

Murad Alaqtash<sup>1</sup>, Huiying Yu<sup>1</sup>, Richard Brower<sup>2</sup>, Amr Abdelgawad<sup>3</sup>,  
Eric Spier<sup>4</sup>, and Thompson Sarkodie-Gyan<sup>1</sup>

<sup>1</sup> Department of Electrical and Computer Engineering, The University of Texas at El Paso,  
El Paso, Texas, 79968, USA

<sup>2</sup> Department of Neurology, Texas Tech University Health Science Center, El Paso,  
Texas, 79905, USA

<sup>3</sup> Department of Orthopedic Surgery & Rehabilitation, Texas Tech University Health  
Science Center, El Paso, Texas, 79905, USA

<sup>4</sup> Mentis Neuro Rehabilitation, El Paso, Texas, 79915, USA

msalaqtash@miners.utep.edu, hyu@miners.utep.edu,  
richard.brower@ttuhsc.edu, amr.abdelgawad@ttuhsc.edu,  
erictspier@hotmail.com, tsarkodi@utep.edu

**Abstract.** The authors have designed and tested a wearable miniature non-invasive sensory system for the acquisition of gait features. The sensors are placed on anatomical segments of the lower limb, and motion data was then acquired in conjunction with electromyography (EMG) for muscle activities, and instrumented treadmill for ground reaction forces (GRF). A relational matrix was established between the limb-segment accelerations and the gait phases. A further relational matrix was established between the EMG data and the gait phases. With these pieces of information, a fuzzy rule-based system was established. This rule-based system depicts the strength of association or interaction between limb-segments accelerations, EMG, and gait phases. The outcome of measurements between the rule-based data and the randomized input data were evaluated using a fuzzy similarity algorithm. This algorithm offers the possibility to perform functional comparisons using different sources of information. It can provide a quantitative assessment of gait function.

**Keywords:** Human Locomotion, Inertial Sensors, Fuzzy Rule-Base, Electromyography.

## 1 Introduction

The predominant method for quantitative human gait analysis is the optical motion system using high-speed cameras to record human 3-D motion [1] and [2]. But there are number of limitations in this laboratory set-up. It requires expensive devices, large space and time-exhausted data processing. Movement of limbs causes occlusion of the reflective markers, requiring reliance on extrapolation by auto-tracking paradigms.

Most importantly, typical optical systems require data processing times that preclude clinical application in real time. Furthermore, the optical motion capture system measures the angular displacements of the anatomical body segments, the measured angles are differentiated twice to obtain the accelerations required to determine the forces. The differentiation introduces unwanted distortions and noise.

An alternative method is using inertial sensors [3 - 6]. The acceleration data can be conveniently collected using wearable accelerometers. It does not require differentiation of data as in optical motion system. This may lead to the possibility of performing gait analysis in real time. Moreover, wearable sensor systems are useful because it is not limited in laboratory, and can be performed for daily applications.

Surface electromyographic (EMG) during gait generates additional important information regarding muscle recruitment patterns and neuromuscular control of walking. Consistent phase-specific patterns of muscle activation occur during gait cycling at preferred walking speeds in normal adults [7] and [8]. The muscles differ from the peak activity at each of the various phases of the gait cycle.

In the analysis of gait, the generation of large amounts of numerical data necessitates aggregation of this data according to their most meaningful patterns and correlations. This aggregation of information makes the partition of space more manageable for further processing. The process of aggregation or granulation implies that we establish the relationships of gait functions, muscle activity patterns, and ground reaction forces. To cope with this issue, a technique is developed using the theory of fuzzy sets.

In this paper, a wearable sensory array was integrated with surface electromyography and instrumented treadmill to perform lower extremity motion analysis, and an intelligent fuzzy computational algorithm was proposed to develop a rule-base (relational matrix) representing muscle activity and acceleration with respect to gait phases throughout the gait cycle.

A fuzzy similarity algorithm was employed to evaluate the similarity between the reference rule-based data and an input test data. This algorithm offers the possibility for obtaining functional comparisons using different sources of information. It can provide a quantitative assessment of gait function.

The purpose of this study was 1) to develop an integration of wearable sensors and devices comprising accelerometers, electromyographic sensors, and instrumented treadmill for ground reaction forces; 2) to develop a relational matrix representing the muscle activity of the lower limb with respect to the gait phases during normal locomotion; 3) to develop a relational matrix representing the acceleration of the lower limb segments with respect to the gait phases; 4) to evaluate and compare the developed rule-based matrices using fuzzy similarity algorithm; and 5) to explore the potential for eventual application in clinical assessment of human gait.

## 2 Methodology

### 2.1 Participants

This study was approved by the Institutional Review Boards (IRB) of the University of Texas at El Paso (UTEP), and all subjects signed an informed consent form prior to

participation. Ten healthy adults were recruited from the campus of UTEP, five male subjects (age  $26.6 \pm 5.5$  years, weight  $79.2 \pm 10$  kg, and height  $181.2 \pm 8.5$  cm) and five female subjects (age  $27 \pm 7.8$  years, weight  $60.6 \pm 9.3$  kg, and height  $162.4 \pm 2.8$  cm). Table 1 illustrates the anthropometric data for all ten healthy subjects.

**Table 1.** Anthropometric data for ten healthy subjects

Gender	Subject	Age (years)	Weight (kg)	Height (cm)	BMI (kg/m <sup>2</sup> )
Male	M1	29	83	186	24.0
	M2	33	80	183	23.9
	M3	29	88	191	24.1
	M4	23	83	177	26.5
	M5	19	62	169	21.7
Mean± std		26.6±5.5	79.2±10	181.2±8.5	24±1.7
Female	F1	21	63	163	23.7
	F2	39	58	161	22.4
	F3	22	66	160	25.8
	F4	31	70	167	25.1
	F5	22	46	161	17.7
Mean± std		27±7.8	60.6±9.3	162.4±2.8	22.9±3.2

## 2.2 Data Acquisition

In this study, the data acquisition system consists of three main hardware devices: instrumented treadmill, surface EMG system, and wearable sensors.

### 2.2.1 Instrumented Treadmill

Each subject was asked to walk bare-footed on an instrumented treadmill (Bertec Corporation, USA) at a self-selected natural speed for three minutes continuously. The self-selected natural speed may be subjective speed selection (comfortable speed) by the subjects. The instrumented treadmill is a dual-belt type with two independent force plates mounted beneath the belts. Each force plate independently measures the GRFs in three directions: X, frontal; Y, sagittal; and Z, coronal. The GRFs data were digitized at 280 Hz. The coronal (vertical) GRF were used to determine the gait phases in a full cycle.

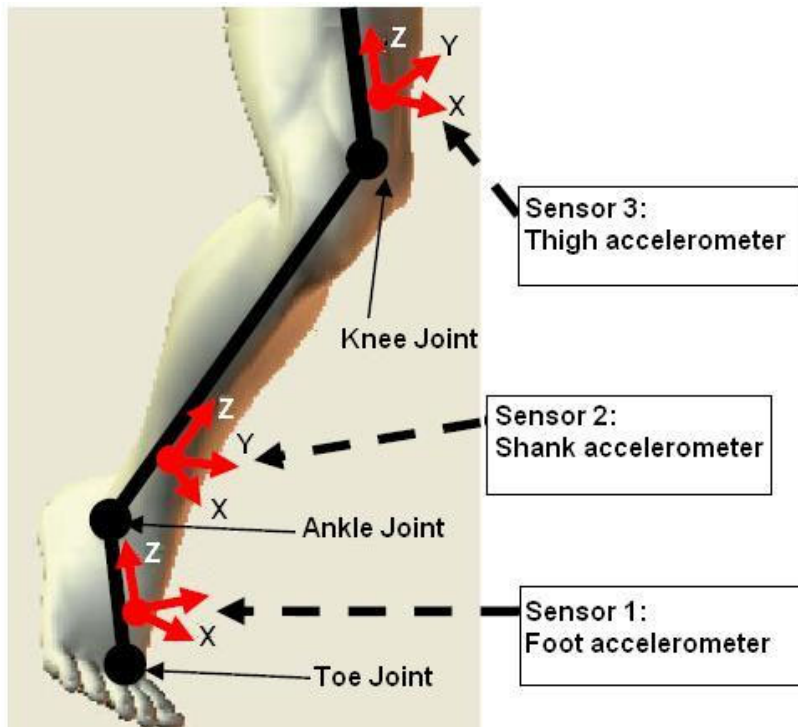
### 2.2.2 Surface EMG

The subjects wore surface EMG electrodes for measuring the dynamic activities of the muscles in both side of the lower extremity. Delsys Myomonitor® wireless EMG system (Delsys Inc., Boston, MA, USA) was used. This system consists of 16 channels, each channel has  $41 \times 20 \times 5$  mm electrode with two  $10.0 \times 1.0$  mm contacts and -92 dB Common Mode Rejection Ratio (CMRR). The electrode placement based on [10] was carefully chosen and the signals were tested before data acquisition. The

acquisition process was performed using Delsys EMGworks® Software at sampling frequency 1000 Hz. Because of the limited number of EMG channels in the system one or two muscles of a muscle group are usually selected to represent the whole group during walking. The following eight muscles for each side were recorded: soleus, tibialis anterior, gastrocnemius, vastus lateralis, rectus femoris, biceps femoris, gluteus medius, and erector spinae.

### 2.2.3 Wearable Sensors

Wearable accelerometers system was developed to detect the acceleration of body segments during walking. The system consists of ADXL330 iMEMS® accelerometer (Analog Devices Inc.). The ADXL330 is a low power and three axis accelerometer with a minimum full-scale range of  $\pm 3g$ . The chip is  $4 \times 4 \times 1.45$  mm and powered by  $V_s=3$  volts. Due to hardware channel limitations, eight accelerometers were used for both left and right legs to record the accelerations in X and Y directions (frontal and sagittal), four on each leg were fitted at the center of mass of foot, shank, thigh and hip. Fig. 1 depicts the placement and orientation of sensors for body segments [5]. The sensors' output was collected at frequency of 280 Hz and stored for processing.



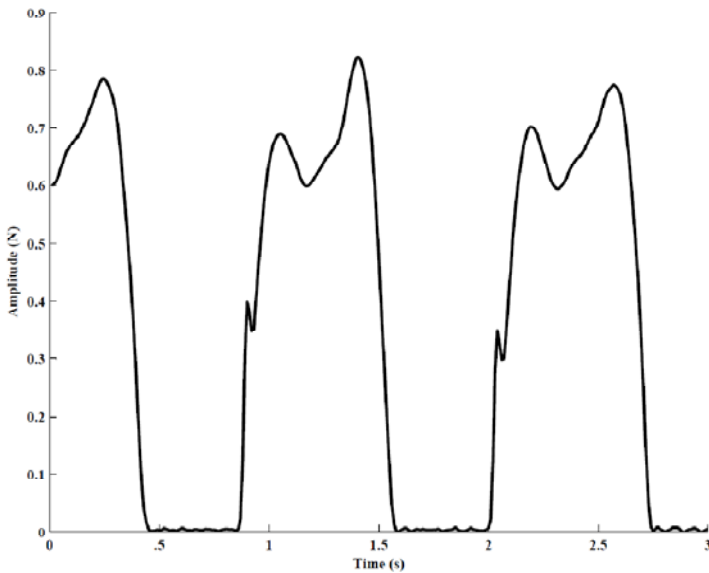
**Fig. 1.** The placement and orientation of sensors for body segments. Adapted from [5].

## 2.3 Data Processing

Different processing techniques were performed based on type of data. In the experiment, three types of raw data were recorded for each subject; ground reaction forces, surface EMG, and accelerations.

### 2.3.1 Ground Reaction Forces

Raw data of GRF were filtered using 2<sup>nd</sup> order Butterworth low pass filter ( $f_c=20$  Hz). Fig 2 shows an example of filtered coronal, vertical, GRF ( $F_z$ ). The start and the end of gait cycle can be determined based on the force  $F_z$ , where the heel strike can be measured based on a selected threshold value equals to 3% of maximum value of  $F_z$ . According to [9], there were seven gait phases in a full gait cycle used in this study; loading response, midstance, terminal stance, preswing, initial swing, midswing, and terminal swing.



**Fig. 2.** Coronal (vertical) GRF  $F_z$  filtered using 2<sup>nd</sup> order Butterworth low pass filter ( $f_c=20$ Hz). X-axis is time (Second) and Y-axis is Amplitude (N).

### 2.3.2 Surface EMG

The processing methods of EMG data shown in Fig. 3 are described in the following steps:

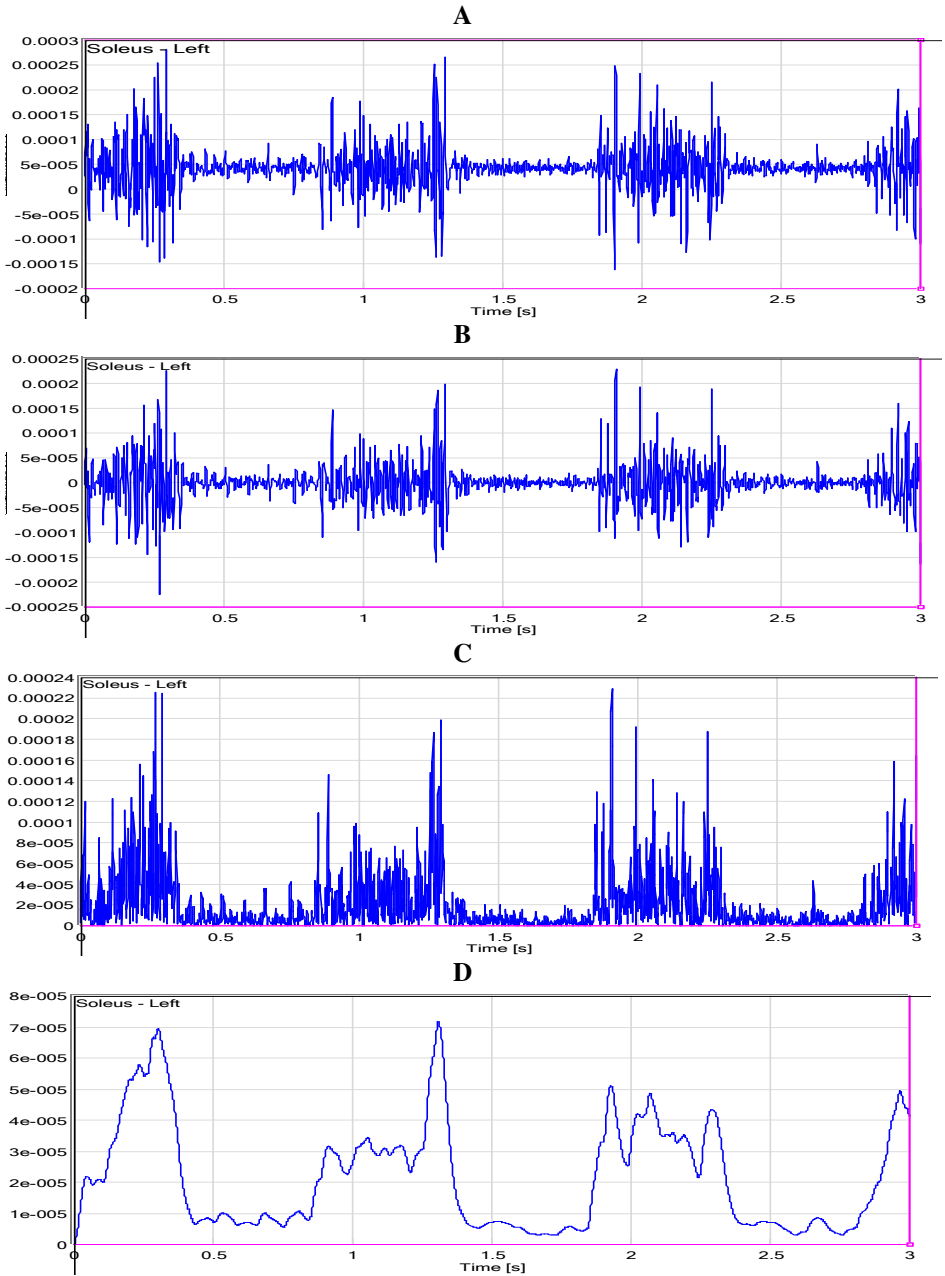
- Step 1. Filtering

Second order Butterworth band pass filter was applied to the raw EMG data between 20 and 200 Hz to reduce the noise effects such as motion artifact noise. Most of the EMG data are concentrated in the band between 20 and 200 Hz [11].

- Step2: Full-wave rectification

The filtered data were then full-wave rectified to generate the absolute value of the EMG.





**Fig. 3.** EMG signal Processing. (A) Row EMG data, (B) 2<sup>nd</sup> order band pass filtered EMG data, (C) Full-wave rectified EMG data, and (D) Linear envelope of EMG data. X-axis is time (Second) and Y-axis is amplitude (Volt).

- Step3: Linear envelope

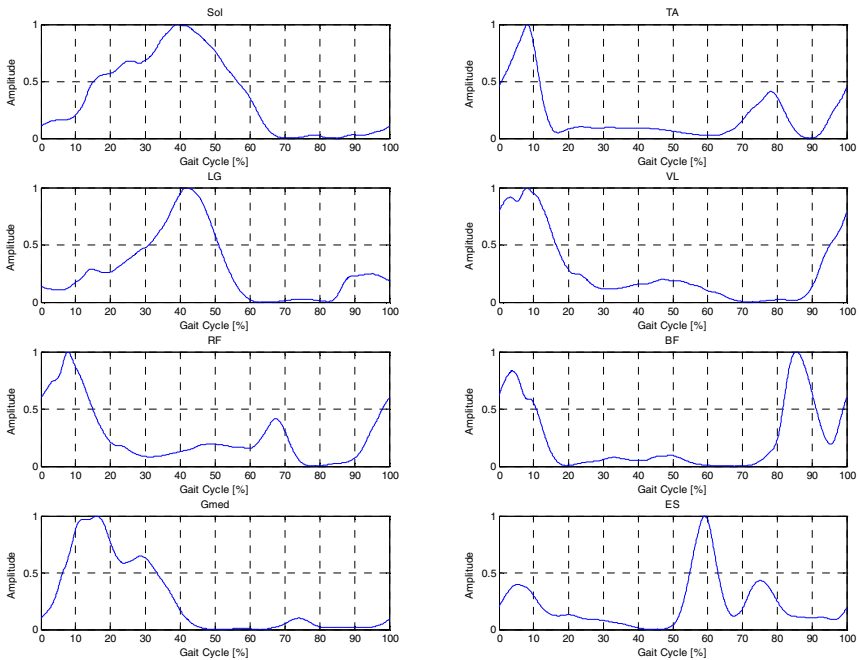
Linear envelope is a common way to manipulate EMG signal [12] and [13]. It can be called also as moving average. The linear envelope was produced by applying a second order Butterworth low pass filter with a cutoff frequency of 7 Hz to the full-wave rectified EMG signal.

- Step4: Averaging over strides

The strides were specified based on the GRF. 100 strides were selected from the three minutes walking. The stride time may vary among strides. Therefore, time-normalization was accomplished by re-sampling and expression each gait cycle in percent cycle rather than time [11].Then, the mean was calculated over all strides.

- Step 5: EMG Normalization

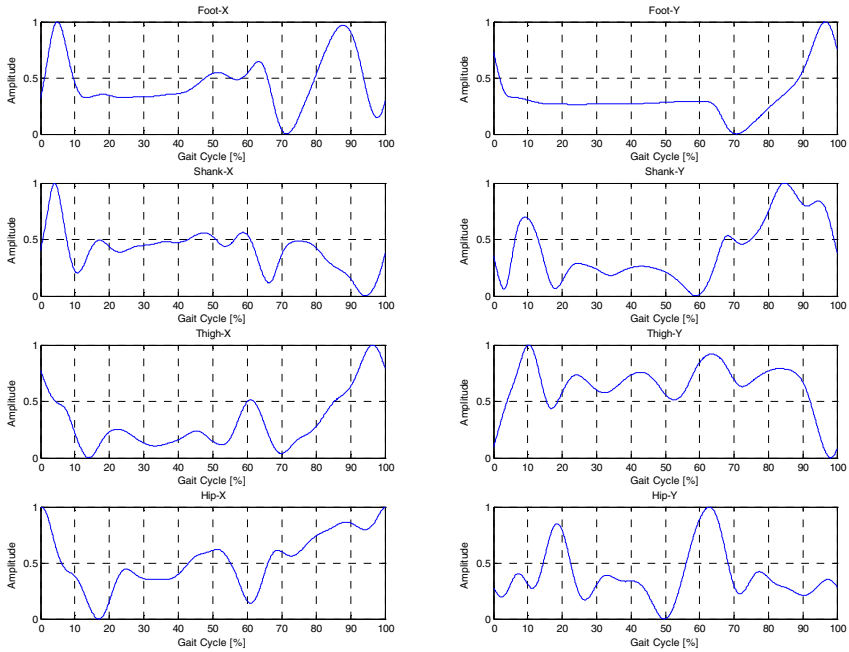
EMG Amplitude normalization was performed because the variability between individuals such as the inherent physiological variability and the variability associated with electrode placement [14]. The normalization was carried out for each subject based on the maximum value of the time-normalized and averaged EMG signal [15]. Fig. 4 shows eight normalized EMG signals of eight muscles for one full gait cycle.



**Fig. 4.** EMG signals of eight muscles for one full gait cycle. X-axis is one full gait cycle and Y-axis is normalized amplitude

### 2.3.3 Accelerations

Second order Butterworth low pass filter with cutoff frequency of 6 Hz was applied to lower the noise and improve the resolution of the accelerometers [11]. As EMG processing procedure, the acceleration data were averaged over 100 strides, followed by normalization based on the maximum acceleration value to get values range from 0 to 1. Fig. 5 depicts the frontal and sagittal accelerations of foot, shank, thigh, and hip for one full gait cycle after processing.



**Fig. 5.** Frontal (X) and sagittal (Y) accelerations of foot, shank, thigh, and hip for one full gait cycle. X-axis is one gait cycle and Y-axis is normalized amplitude

## 2.4 Formation of Fuzzy Rule-Based Relational Matrix

### 2.4.1 Fuzzy Sets

In this study, three fuzzy sets were defined to represent muscle activities, gait phases, and accelerations.

The fuzzy set of the EMG data was described as the muscle type  $x$  from a muscle group  $X$ , expressed as  $x \in X$ ,  $X = \{Sol \ TA \ LG \ VL \ RF \ BF \ Gmed \ ES\}$  where  $\{\dots\}$  indicates muscles and *Sol* – soleus; *TA* – tibialis anterior; *LG* – gastrocnemius lateralis; *VL* – vastus lateralis; *RF* – rectus femoris; *BF* – biceps femoris; *Gmed* – gluteus medius; *ES* – erector spinae.

The fuzzy set of the seven gait phases was depicted as the gait phase  $y$  out of a number of gait phases  $Y$ , expressed as  $y \in Y$ ,  $Y = \{Phase1 \ Phase2 \ Phase3 \ Phase4 \ Phase5 \ Phase6 \ Phase7\}$ ; where  $\{\dots\}$  indicates the seven gait phases and *Phase1* is

loading response; *Phase2* is midstance; *Phase3* is terminal stance; *Phase4* is preswing; *Phase5* is initial swing; *Phase6* is midswing; *Phase7* is terminal swing.

The third fuzzy set was described as the acceleration  $z$  from the acceleration group  $Z$ , expressed as  $z \in Z, Z = \{Foot_x, Foot_y, Shank_x, Shank_y, Thigh_x, Thigh_y, Hip_x, Hip_y\}$ , where  $\{\dots\}$  indicates the frontal and sagittal accelerations of foot, shank, thigh, and hip.

In fuzzy set theory, a member of a set is represented by a degree between 0 and 1. This degree is called membership degree which shows belonging degree of the member to the class.

### 2.4.2 Fuzzy Relations

The association and interaction between the elements of muscle activity, gait phases, and accelerations are represented by fuzzy relational matrices.

Fuzzy relational matrix can be obtained for each two fuzzy sets using the membership function described as:

$$P(x, y) = \mu_p(x, y) = \frac{1}{n} \sum_{i=1}^n \mu(i) \tag{1}$$

Where  $\mu_p(x, y)$  the membership function,  $n$  is the number of data frames in a fuzzy set, and  $\mu(i)$  is the normalized data of that fuzzy set from 0 to 1. For instance,  $\mu_p(Sol, Phase1)$ , is the membership degree of *Soleus* muscle within the loading response phase.

Two fuzzy relational matrices were defined. The first relational matrix was defined as  $P(x,y)$ , the relation between EMG data and gait phases; and the second relational matrix was defined as  $Q(z,y)$ , the relationship between gait phases and accelerations.

The developed fuzzy relational matrices depict the rule-based system that describes the strength of association and interaction between the elements of EMG, gait phases, and accelerations.

### 2.4.3 Fuzzy Similarity

This algorithm offers a comparison between the reference rule-base and the acquired test data during gait cycle, and it is expressed as the fuzzy similarity algorithm. It provides an evaluation algorithm to determine the behavior of muscle activities and limb-segments accelerations in the corresponding gait phases.

The fuzzy similarity measure is depicted by:

$$\begin{aligned} \mu_{ref} * \mu_{test} &= \frac{\mu_{ref} \wedge \mu_{test}}{\mu_{ref} \vee \mu_{test}} \\ &= \frac{\min \{ \mu_{ref}(x, y), \mu_{test}(x, y) \}}{\max \{ \mu_{ref}(x, y), \mu_{test}(x, y) \}} \end{aligned} \tag{2}$$

Where  $*$  represents the fuzzy similarity operator,  $\mu_{ref}$  is the membership function of the reference subjects,  $\mu_{test}$  is the membership function of the test subjects,  $\wedge$  depicts the minimum, and  $\vee$  depicts the maximum. The resulting similarity between the reference and test attributes of the subjects, expressed through the grade of similarity range between 0 and 1.

### 3 Results and Discussions

Five male and five female subjects were recruited in this experiment and were analyzed separately due to anatomical differences especially with respect to the muscles [16], bones, and skeletal morphology. Three of the subjects were used for training to establish the reference fuzzy rule-base, and the other two were used as randomized subjects for testing.

The reference rule-based matrices, therefore, are expressed as follows:

$$P_{ref}(x, y) = \begin{bmatrix} .295 & .588 & .889 & .628 & .226 & .200 & .155 \\ .871 & .254 & .140 & .084 & .337 & .524 & .312 \\ .250 & .438 & .797 & .416 & .144 & .156 & .197 \\ .904 & .476 & .230 & .329 & .354 & .205 & .495 \\ .681 & .541 & .420 & .432 & .590 & .307 & .295 \\ .655 & .385 & .308 & .164 & .145 & .260 & .621 \\ .594 & .655 & .317 & .150 & .162 & .179 & .194 \\ .558 & .278 & .183 & .697 & .501 & .441 & .413 \end{bmatrix}$$

$$Q_{ref}(y, z) = \begin{bmatrix} .674 & .442 & .662 & .477 & .523 & .591 & .596 & .303 \\ .462 & .297 & .695 & .419 & .116 & .728 & .188 & .420 \\ .556 & .302 & .705 & .305 & .197 & .667 & .311 & .265 \\ .701 & .312 & .726 & .120 & .440 & .499 & .336 & .372 \\ .581 & .192 & .765 & .421 & .429 & .789 & .609 & .781 \\ .472 & .171 & .901 & .697 & .424 & .782 & .801 & .288 \\ .656 & .767 & .313 & .843 & .830 & .465 & .923 & .216 \end{bmatrix}$$

Where  $P_{ref}(x,y)$  depicts the strength of relation between the muscles activities and the gait phases.  $Q_{ref}(y,z)$  depicts the strength of the relation between the gait phases and the limb-segment accelerations.

The rule-based system, relational matrices, elements represent the membership degree of individual muscle activity and segment acceleration in a particular gait phase within gait cycle. For instance,  $\mu_p(Sol, Phase3)$ , depicts high activity ( $P_{ref}(1,3)= 0.889$ ) of *Soleus* muscle within the terminal stance phase. While the same muscle has low activity ( $P_{ref}(1,7)= 0.155$ ) within the terminal swing phase as depicted by  $\mu_p(Sol, Phase7)$ . This algorithm provides an evaluation methodology for determining the behavior of each individual muscle or body segment in a corresponding gait phase.

A comparison of the muscle activities depicts a grade of similarity 0.78 for male and 0.82 for female between the reference rule-based matrix,  $P_{ref}$ , and test subjects matrix,  $P_{test}$ , from the reference. Similar comparison for the randomized test subjects depicts 0.71 for male and 0.73 for female. The average similarities were evaluated at 0.75 and 0.78 for male and female respectively.

Another comparison for the body segments accelerations between the rule-based matrix,  $Q_{ref}$ , and test subjects matrix,  $Q_{test}$ , exhibits average similarities of 0.82 for male and 0.81 for female. The fuzzy similarity measure offers the possibility for obtaining functional comparisons using different sources of information.

**Table 2.** The Average Similarity

	Male		Female	
	$P_{ref} * P_{test}$	$Q_{ref} * Q_{test}$	$P_{ref} * P_{test}$	$Q_{ref} * Q_{test}$
Reference subjects	0.78	0.83	0.82	0.85
Randomized subjects	0.71	0.80	0.73	0.77
Average	0.75	0.82	0.78	0.81

## 4 Conclusion

This algorithm has clearly illustrated the possibility to perform functional comparisons by using different sources of information. The fuzzy similarity methodology depicts distinctions between the reference able-bodied and the randomized test subjects within with a membership grade of belonging.

This novel algorithm may offer very reliable and efficient tools for the evaluation and assessment of gait function in several ways:

1. By building a rule-based system depicts the strength of relation between muscle activities, limb-segment accelerations, and gait phases.
2. By comparing the reference muscle activities within gait phases with a randomized input-matrix through a fuzzy similarity algorithm,
3. By comparing the reference limb-segment accelerations within gait phases with a randomized input-matrix through a fuzzy similarity algorithm,

**Acknowledgments.** The authors wish to thank the Stern Foundation for providing the funds for this research work. The authors also would like to thank all the subjects who participated in this study.

## References

1. Karaulovaa, I.A., Hallb, P.M., Marshall, A.D.: Tracking people in three dimensions using a hierarchical model of dynamics. *Image and Vision Computing* 20, 691–700 (2002)
2. Gavril, D.M., Davis, L.S.: 3-D model-based tracking of humans in action: a multi-view approach. In: *IEEE Computer Vision and Pattern Recognition*, San Francisco, pp. 73–79 (1996)
3. Kavanagh, J.J., Menz, H.B.: Accelerometry: a technique for quantifying movement patterns during walking. *Gait and Posture* 28(1), 1–15 (2008)
4. Takeda, R., Tadano, S., Todoh, M., Yoshinari, S.: Human Gait Analysis using Wearable Sensors of Acceleration and Angular Velocity. In: *13th Int. Conf. on Biomed. Eng.*, vol. 23, pp. 1069–1072 (2009)
5. Tao, L., Yoshio, I., Kyoko, S., Haruhiko, K.: Development of a wearable sensor system for quantitative gait analysis. *Measurement* 42(7) (2009)
6. Mayagoitaia, R.E., Nene, A.V., Veltink, P.H.: Accelerometer and rate gyroscope measurement of kinematics: an inexpensive alternative to optimal motion analysis systems. *Biomechanics* 35(4), 537–542 (2002)

7. Arsenault, A.B., Winter, D.A., Marteniuk, R.G., Hayes, K.C.: How many strides are required for the analysis of electromyographic data in gait? *Scand. J. Rehabil. Med.* 18, 133–135 (2001)
8. Winter, D.A.: *Biomechanics and Motor Control of Human Gait: Normal, Elderly and Pathological*. Waterloo Biomechanics Press, Waterloo (1991)
9. Perry, J.: *Gait Analysis: Normal and Pathological Function*, pp. 11–15. SLACK Inc., Thorofare (1992)
10. Cram, J.R., Kasman, G.S., Holtz, J.: *Introduction to Surface Electromyography*. Aspen Publishers, Maryland (1998)
11. Winter, D.A.: *Biomechanics and Motor control of Human Movement*, 4th edn. John Wiley & Sons, Chichester (2009)
12. De Luca, C.J.: The use of surface electromyography in biomechanics. *J. Applied Biomechanics* 13(4), 135–163 (1997)
13. De Luca, C.J.: *Electromyography Encyclopedia of Medical Devices and Instrumentation*, pp. 98–109. John Wiley, Chichester (2006)
14. Allison, G.T., Marshal, R.N., Singer, K.P.: EMG Signal Amplitude Normalization Technique in Stretch-shortening Cycle Movements. *J. Electromyography and Kinesiology* 3(4), 236–244 (1993)
15. Ricamato, A.L., Hidler, J.M.: Quantification of the dynamic properties of EMG patterns during gait. *J. Electromyography and Kinesiology* 15, 384–392 (2005)
16. Pincivero, D.M., Campy, R.M., Salfetnikov, Y., Bright, A., Coelho, A.J.: Influence of contraction intensity, muscle, and gender on median frequency of the quadriceps femoris. *J. Appl. Physiol.* 90(3), 804–810 (2001)

# The Mathematical Model and Control Scheme of a Four-Legged Robot Based on GZ-I and Note Module

Yinfeng Fang<sup>1</sup>, Houxiang Zhang<sup>2</sup>, Xiuli Li<sup>1</sup>, and S.Y. Chen<sup>1,\*</sup>

<sup>1</sup> Institute of Intelligent Systems, Zhejiang University of Technology  
310023 Hangzhou, China

<sup>2</sup> Institute of TAMS, Department of Informatics, University of Hamburg  
Vogt-Koelln-Strasse 30, D-22527 Hamburg, Germany  
sy@ieee.org

**Abstract.** This paper adopts a set of GZ-I robot modules and a Note module for construction of an H-shaped four-legged robot. Kinematics graph theory and homogeneous matrix transform are applied for gait design and locomotion planning. Its forward crawling motion is designed and simulated by computers. The corresponding control scheme of the modular robot is also studied in this paper, which can control 8-channel modules and realizes rotational speed adjustable by an 8-bit microcontroller. Experiments show the feasibility of the system structure and effectiveness of the motion control.

**Keywords:** Modular robots, GZ-I module, four-legged robot, kinematics graph theory, homogeneous matrix transform, PWM, locomotion.

## 1 Introduction

Modular robots are made up of small, independent components that can be replicated and combined in different ways to perform various tasks. Each module can be manufactured in large numbers, reducing production costs. Defective modules can be quickly replaced, and can be designed to detach and reassemble rather than break when they are struck [1]. There have appeared many modular robots, such as M-TRAN [2][3], PolyBot [4][5], Superbot[6], CONRO [7], YaMoR [8], Molecube [9], and so on.

In this paper, we adopted a modular robot prototype GZ-I project which is based on the cooperation with Juan González-Gómez from the School of Engineering, Universidad Autonoma de Madrid in Spain and Houxiang Zhang from TAMS group in University of Hamburg, Germany [10]. The GZ-I module has three connected surfaces for attaching other modules, so the modules can mainly been constructed to footless robots, like 1D pitch and pitch-yaw connecting modular robots that have been presented in previous papers [10][11][12][13]. Besides the linear structures, Yong Li from the Zhejiang University of Technology designed a four-legged robot based on GZ-I module [14]. Besides the GZ-I module, we also adopted a Note module that

---

\* Corresponding author.



designed by our research group. The Note module is used to connect with GZ-I module, and construct a more symmetrical robot.

In the field of modular robot, Central Pattern Generator (CPG) plays an important role in expressing motion. Almost all modular robots adopt CPG model. Department of Brain Science and Engineering at Kyushu Institute of Technology introduced the Matsuoka Model as the neuron model of CPG to express the motion of a snake-like modular robot [15]. School of Computer and Communication Science, Ecole Polytechnique Fédérale de Lausanne used CPG and a gradient-free optimization algorithm to complete the task of structuring different robotic structures, and learning to locomote [16]. National Institute of Advanced Industrial Science and Technology at Tokyo designed several locomotion patterns in various structures using a CPG controller model and GA optimization [17]. Department Physics and Electronic Engineering, YunNan Nationalities University present a CPG model for machine crab's octopod gait [18]. Nonlinear Systems Laboratory at Cambridge simplified modular architectures based on CPGs for robotic applications, and show their global exponential stability using partial contraction analysis [19].

CPGs underlie the production of most rhythmic motor patterns and have been extensively researched as models of neural network function. However, CPG model is not regarded as the specific position of the related connector, but oscillator's rhythmic rotation, let alone the other modular's. They get CPG parameters through complex algorithm. So Kinematics Graph Theory is introduced thoroughly so as to set up robot's model geometry, and gives each module one or two 3-D coordinates. Furthermore, homogeneous matrix transform as quoted describe the module's rotation. Base on these, we try to find CPG parameters easily and simulate the crawling locomotion in Matlab according to the gait described in article [14].

Here it is important constituent of quoting a way of GZ-I mass-controls, especially generating 8-channel PWMs based on 8-bit microcontroller. This way supplies a parameter, which is a critical factor of rotational speed of GZ-I. Adjustable rotational speed is indispensable for obtaining a stationary locomotion.

The rest of the paper is organized as follows. Section 2 describes the GZ-I module and the Note module. Section 3 presents the design of the four-legged robot and the mathematical model. Section 4 presents the method to control 8-channel module with speed adjustable. Section 5 describes the result of experiments. Section 6 concludes the paper with future work.

## 2 GZ-I Modules and Note Modules

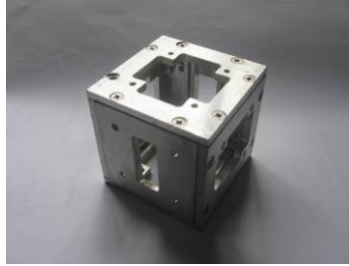
### 2.1 GZ-I Module

GZ-I module is composed of RC servo and module framework. It has only one degree of freedom which is actuated by an RC servo. The rotation range is 180 degrees. This module has three junction surfaces: two of them at the bottom and the rest on the side.

Using these surfaces, pitch-pitch structure robot, pitch-yaw structure one, and simple four-legged one can be made up. These kinds of robots and their locomotion have been mentioned in other scholars' paper [1-4], [14].

## 2.2 Note Module

Note module is composed of 6 pieces of connecting surfaces, and each surface can connect Note module with GZ-I module. Figure 1 shows the Note Module. This module is designed for constructing much more complicated robot and better performance.

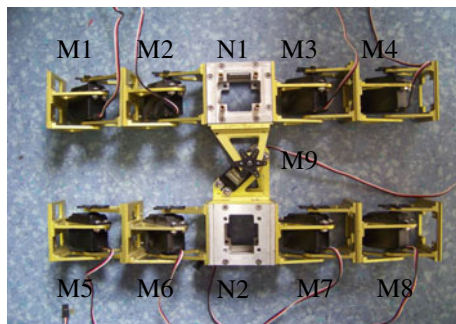


**Fig. 1.** Note Module

## 3 Four-Legged Robot and Mathematical Model

### 3.1 Structural Design

Based on these GZ-I modules and Note modules, an H-shaped four-legged robot constructed here. It consists of nine GZ-I modules and two Note modules. Eight GZ-I modules rotate around pitch axis, and the middle GZ-I module rotates around the yaw axis. Two Note modules are used to connect with GZ-I modules. Each leg consists of two lateral GZ-I modules in the pitch-pitch way, and the waist is the transverse one. The structure consisted by modules is symmetrical that is able to implement a crawling locomotion conveniently. Figure 2 shows an example of the four-legged robot.



**Fig. 2.** M1 and M2 are joined up by pitch-pitch way, making up robot's up-left leg. M3 and M4 are robot's up-right leg. M5 and M6 are robot's down-left. M7 and M8 are robot's down-right leg. M9 is the waist. N1, N2 are just for connecting each part.

### 3.2 Mathematical Model

It is crucial to build a mathematical model for describing the structure of the above four-legged robot in a reasonable way. The robot's state and locomotion can be demonstrated in mathematics. In this paper, kinematics graph theory is brought in to take a mathematical model for the robot. In terms of this theory, the robot consists of joints and connecting rods. The former can revolve around an axis, while the latter can't. Two rods can be only bonded with one joint, and one using for over two. With view to explain these characteristics simply, joints and rods are shown by lines and dots respectively.

As the above theory mentioned, the initial state of the four-legged robot is painted in Matlab. Each leg is composed of three rods and two joints. The waist is made up of one joint and two connecting rods. Obviously, two connecting rods are shared by three joints.

According to the robot's physical structure exactly, the following principle is a best way to choose coordinates. For example, the coordinate of Vc4 is the endpoint of the rod and Vb4's is midpoint of the rod.

There are two homogeneous matrixes to describe robot's posture, which are rods matrix R and joints matrix J.

The homogeneous matrix of rods can be expressed as below as robot's initial state.

$$\mathbf{R} = \begin{bmatrix} Va1 \\ Va2 \\ Vb1 \\ Vb2 \\ Vb3 \\ Vb4 \\ Vc1 \\ Vc2 \\ Vc3 \\ Vc4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & d+s & 0 & 1 \\ -(s+d/2) & 0 & 0 & 1 \\ s+d/2 & 0 & 0 & 1 \\ s+d/2 & d+s & 0 & 1 \\ -(s+d/2) & d+s & 0 & 1 \\ -(2s+d/2) & 0 & 0 & 1 \\ 2s+d/2 & 0 & 0 & 1 \\ 2s+d/2 & d+s & 0 & 1 \\ -(2s+d/2) & d+s & 0 & 1 \end{bmatrix},$$

where  $d$  is length of a side at the bottom of one piece, and  $s$  is GZ-I module's length. Figure 3 shows dimensions of GZ-I module.

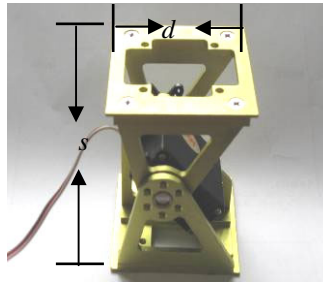


Fig. 3. Dimensions of GZ-I module

The other homogeneous matrix **J** is as below.

$$\mathbf{J} = \begin{matrix} \left[ \begin{matrix} La \\ Lb1 \\ Lb2 \\ Lb3 \\ Lb4 \\ Lc1 \\ Lc2 \\ Lc3 \\ Lc4 \end{matrix} \right] \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix} = \begin{bmatrix} 0 & (d+s)/2 & 0 & 1 \\ -(d+s)/2 & 0 & 0 & 1 \\ (d+s)/2 & 0 & 0 & 1 \\ (d+s)/2 & d+s & 0 & 1 \\ -(d+s)/2 & d+s & 0 & 1 \\ -(d+s)/2-s & 0 & 0 & 1 \\ (d+s)/2+s & 0 & 0 & 1 \\ (d+s)/2+s & d+s & 0 & 1 \\ -(d+s)/2-s & d+s & 0 & 1 \end{bmatrix}$$

The matrix **J** is much more important than **R**, because each joint revolve around an axis in **J**. Here, all joints are classified into two following specific types: horizontal joint, which can rotate around horizontal axis, such as La; aerial joint, which can rotate by vertical axis, such as Lb1, Lb2, Lb3, Lb4, Lc1, Lc2, Lc3, Lc4.

### 3.3 Joint Rotating

When to consider joint rotating is a key point. As mentioned, one joint is connected with two rods. So when one joint spins, it will turn up two kinds of cases. One case is both rods changed, and the other is one changed and the other rod not. To solve the uncertainty, this paper stipulates two rules depended on the practical operation of a robot. One is for horizontal joint, and the other for aerial joint.

**Rule 1** is described as below. Each horizontal joint belongs to a coincident leg. When one horizontal joint spins, the rod near the end of leg changes position but the rod near the waist does not. La1, Lb1, Lb2, Lb3, Lb4, Lc1, Lc2, Lc3, Lc4 should comply with **Rule 1**.

**Rule 2** is applied for La, because La is the only aerial joint. Obviously, the robot is described in part A, which is completely symmetrical with the central La1. So rule 2 supposes that two rods connected with La1 should be changed together in a symmetrical form. This rule has been proved to be correct in experiment, ignoring other factors, such as the center of gravity and friction.

Complied with two rules stipulated above, robot's state is unique after any joint spins a certain angle. Thus robot's changes caused by joint spinning can be expressed by homogeneous matrix transform. Furthermore robot's locomotion gaits can be described by matrix transform. This work supplies an expedient way to gain an effective and stabilize gait and a convenient approach to get robot's movement performance, such as walking speed and obstacle-climbing capability.

Combining the rules and homogeneous matrix transform, robot's standing state and forward crawling gait has been obtained using Matlab.

## 4 The Control System for Robot

The aim of control system is the design of an interface between microcontroller and GZ-I Modules. Microcontroller generate pulse width modulation (PWM) signal to control masses of GZ-I modules and implement robot's gait, such as forward crawling and turning.

### 4.1 The Control Signal of GZ-I Module

GZ-I module is a single-degree-of-freedom module. The motive power is supplied by an RC servo motor. It is the same to control the servo motor as the GZ-I module. RC servo motor has three lead wires, power, ground and PWM. This paper is only concerned with the PWM lead wires. The period and the pulse width of PWM are fixed in factory. Period is 20ms, and pulse width changes from 1.5ms to 2.5ms mapping to the angular position of servo motor from -90 degrees to 90 degrees.

The expression between angular position and the pulse width of PWM is

$$d = 90(t-1.5) \quad 0.5\text{ms} \leq t \leq 2.5\text{ms}, \quad (1)$$

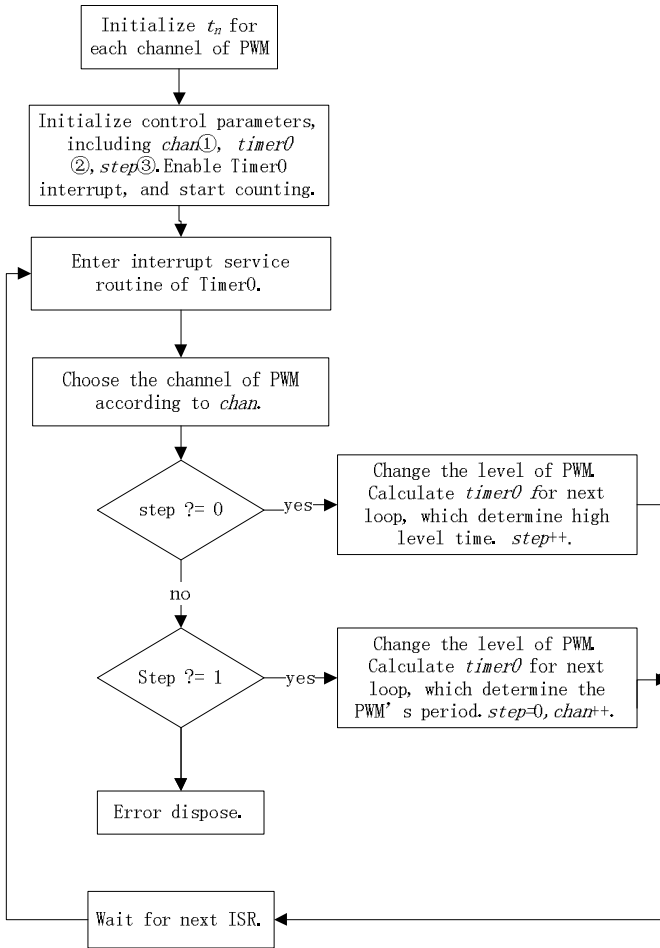
where  $d$  is the angular position of servo motor, and  $t$  is pulse width.

The target of our design is to make program to meet the following two demands. A) Eight GZ-I modules should be controlled at the same time, or in a relatively short time. Only in this way, can all modules rotate to form a gait continuously. B) To reduce impact that comes from the colliding between robot's leg and ground, and to implement modules' synchronous movement, rotational speed should be adjustable.

### 4.2 Generating 8-Channel PWM for Robot

Methods to generate multichannel PWMs are abundant. This article adopt microcontroller to do it because of convenience and efficiency. It is suitable for modular robot control. Choosing suitable parameters for timer that is integrated in microcontroller, microcontroller would output satisfied PWM through IO interface. The kernel idea of generating 8-channel PWM is that microcontroller allocate time slice that is 2.5ms in our system to each channel of PWM, and the sum of time slices is right 20ms, which is the period of PWM. During every time slice, the corresponding channel should be changed output level twice to get a pulse width. First change low to high. Then keep the high level for a while ranged from 0.5 ms to 2.5 ms, which determine the angle of servo motor. At last turns it back to low level.

The process proposed above has been realized in our program. Figure 4 shows the algorithm framework.



**Fig. 4.** ① is used to choose the corresponding PWM channel. ② is used to set TH0 and TL0, which is related to T, t, t<sub>1</sub>,..., t<sub>n</sub>. ③ is used to choose corresponding service program for each channel of PWM.

### 4.3 Implementing Rotational Speed Adjustable

A method about adjusting servo motor's rotate speed is proposed here. In order to facilitate the procedure, we use only one parameter to adjust rotate speed. When servo motor receives a PWM signal with a constant pulse width, the motor stops at a corresponding angular position after rotated a piece of time at a maximum practical speed 0.2 s per 60 degrees. To slow down and to adjust the speed, this article brings in a parameter *p*, which split the whole rotational step to short steps. Different rotational speed can be obtained by changing the value of *p*. The rotational speed can be worked out as the formula below

$$v = \frac{p \cdot T}{\Delta\beta}, \quad (2)$$

where  $T$  is the periods of PWM, and  $\Delta\beta$  is the difference value of start angle and end angle.

## 5 Experimental Results

The experiments are mainly to test the methods mentioned in section 3 and section 4. The crawling locomotion gait have been successfully implemented and tested on the robot. The experiment measured the velocity of crawling locomotion and compared it with theoretical value got in Matlab.

Because of the bad performance without rotational speed adjusting, we failed to complete the forward crawling before. Through observation it is due to enormous impact from the ground and joints rotating out of sync. Thus this paper found the way of adjusting rotational speed to solve the problem. Obviously, lower rotate speed can reduce the impact and the same parameter  $p$  for each channel can achieve synchronization.

Velocity is an important property to robot. So this experiment measured the velocity of crawling locomotion and compared it with theoretical value got in Matlab. Set the amplitude of  $L_a$  to be 25 degrees, which determine the stride length and taking  $p$  different values, and experiments yielded different theoretical velocities (TV) and practical velocities (PV) as showed in table.1 below.

**Table 1.** Different TV and PV with different parameter  $p$

$p$	10	15	20	30	50	80
TV(cm/s)	15.5	11.6	7.8	5.2	3.1	1.9
PV(cm/s)	5.9	5.4	4.6	3.1	1.8	1.2

It shows that when  $p$  is larger than 30, there is little difference between PV and TV. The main reason for the error probably is the simplified model and mechanical friction. When  $p$  is less than 20, the error becomes larger and larger. Through observing the experiment of crawling, we found out robot's feet skidded on the ground severely when  $p$  is less then 20. This should be the reason why the velocity is not ascent obviously with the decreasing of  $p$ . With the experiment result Fig 11, it is easy to control robot's velocity by choosing different  $p$  in the PV line. It must be noted that the velocity is restricted in the range from 0 cm/s to 6 cm/s. And increasing friction coefficient may be a good way to expand scope of speed control.

In the experiments, we unavoidably met some mistakes in the system. One mistake is that the robot can't go straight in a long distance. It is probably caused by the asymmetry of machine structure and the uncertainties caused by coupling, friction and payload masses. The structure has to be improved in our future work.

## 6 Conclusion and Future Work

Based on GZ-I module and Note module, this paper designed a four-legged robot, and designed a system to complete the crawling gait. In this system, controlling masses of servo motors and achieving speed adjustable were proposed. This paper also introduced Kinematics Graph Theory to describe robot's mechanical structure, and homogeneous matrix is used to build the mathematical model. In addition, two formulas were proposed to describe joints rotating. Crawling gaits were simulated in Matlab.

However, this paper obtained robot's CPG parameters that is not optimal using manual mode, and can not reflect the superiority of mathematical model. Our next work is to obtain CPG parameters with learning or optimizing algorithm by the assist of mathematical model to get more efficient gaits, such as side crawling and turning. Besides, we are also attempting to add some sensors to make the robot more intelligent.

## References

1. Roberts, G.: A Modular Robot That Puts Itself Back Together Again, vol. 3. New York Times (July 28, 2009)
2. Murata, S., Yoshida, E., Kamimura, A., Kurokawa, H., Tomita, K., Kokaji, S.: M-TRAN: Self-Reconfigurable Modular Robotic System. *IEEE/ASME Transactions on Mechatronics* 7(4), 431–441 (2002)
3. Kurokawa, H., Kamimura, A., Yoshida, E., Tomita, K., Murata, S., Kokaji, S.: Self-Reconfigurable Modular Robot (M-TRAN) and its Motion Design. In: *Seventh International Conference on Control, Automation, Robotics And Vision*. Singapore (2002)
4. Yim, M., Zhang, Y., Duff, D.: Modular Robots. *IEEE Spectrum Magazine*, 30–34 (February 2002)
5. Yim, M., Roufas, K., Duff, D., Zhang, Y., Eldershaw, C., Homans, S.: Modular Reconfigurable Robots in Space Applications. *Autonomous Robot Journal*. Special Issue for Robots in Space (2003)
6. Shen, W.M., Krivokon, M., Chiu, H., Everist, J., Rubenstein, M., Venkatesh, J.: Multi-mode locomotion via SuperBot reconfigurable robots. *Autonomous Robots* 20(2), 165–177 (2006)
7. Castano, A., Shen, W.M., Will, P.: CONRO: Towards Deployable Robots with Inter-Robot Metamorphic Capabilities. *Autonomous Robots* 8(4), 309–324 (2000)
8. Moeckel, R., Jaquier, C., Drapel, K., Dittrich, E., Upegui, A., Ijspeert, A., Yamor: Bluemove.: An autonomous modular robot with Bluetooth interface for exploring adaptive locomotion. In: *Proceeding of the 8th International Conference on Climbing and Walking Robots*, London, U.K, pp. 685–692 (2005)
9. Zykov, V., et al.: Evolved and Designed Self-Reproducing Modular Robotics. *IEEE Transactions on Robotics* 23(2), 308–319 (2007)
10. Zhang, H., Gonzalez-Gomez, J., Xie, Z., Cheng, S., Zhang, J.: Development of a Low-cost Flexible Modular Robot GZ-I. In: *Proc. of 2008 IEEE/ASME Intl. Conf. on Advanced Intelligent Mechatronics*, Xi'an, China, pp. 4–7 (2008)
11. Gonzalez-Gomez, J., Boemo, E.: Motion of Minimal Configurations of a Modular Robot: Sinusoidal, Lateral Rolling and Lateral Shift. In: *Proc. of the 8th Intl. Conf. on Climbing and Walking Robots*, London, UK, pp. 667–674 (2005)



12. Gonzalez-Gomez, J., Zhang, H., Boemo, E., Zhang, J.: Locomotion Capabilities of a Modular Robot with Eight Pitch-Yaw-Connecting Modules. In: The 9th Intl. Conf. on Climbing and Walking Robots and their Supporting Technologies for Mobile Machines, Brussels, Belgium, pp. 12–16 (2006)
13. Gonzalez-Gomez, J., Zhang, H., Boemo, E.: Locomotion Principles of 1D Topology Pitch and Pitch-Yaw-Connecting Modular Robots. In: Bioinspiration and Robotics: Walking and Climbing Robots, Vienna, Austria, pp. 403–428 (2007)
14. Li, Y., Zhang, H., Chen, S.Y.: A four-legged robot based on GZ-I modules. In: 2008 IEEE International Conference on Robotics and Biomimetics, Bangkok, Thailand, pp. 921–926 (2009)
15. Matsuo, T., Kazuo, I.: A CPG control system for a modular type mobile robot. International Congress Series, vol. 1301, pp. 206–209 (2007)
16. Sproewitz, A., et al.: Learning to Move in Modular Robots using Central Pattern Generators and Online Optimization. International Journal of Robotics Research 27(3/4), 423–443(2008)
17. Kurokawa, et al.: Self-reconfigurable M-TRAN structures and walker generation. Robotics & Autonomous Systems 54(2), 142–149 (2006)
18. He, J., Lu, C.F., Yin, S.T.: The design of CPG control module of the bionic mechanical crab. In: 2006 IEEE International Conference on Robotics and Biomimetics, ROBIO 2006, pp. 280–285 (2006)
19. Seo, K., Slotine, J.J.: Models for global synchronization in CPG-based locomotion. In: Proceedings - IEEE International Conference on Robotics and Automation, pp. 281–286 (2007)

# Electromechanical Coupling Model for Cutterhead Driving System of Shield Machines

Jianzhong Sun<sup>1</sup>, Ran Liu<sup>1</sup>, Yaqin Luo<sup>1</sup>, and Wei Sun<sup>2</sup>

<sup>1</sup> School of Electrical Engineering, Dalian University of Technology,  
Dalian 116023, Liaoning Province, China

<sup>2</sup> School of Mechanical Engineering, Dalian University of Technology,  
Dalian 116023, Liaoning Province, China

jzsun@dlut.edu.cn, psincos@163.com, sunwei@dlut.edu.cn

**Abstract.** The synchronization of the cutterhead driving system of shield machines is affected by not only the nonlinearity of mechanical transmission mechanism and the characteristics of driving motors, but also their interaction. In this paper, dynamics of the motor-driving system of the cutterhead is analyzed, and the interaction mechanism between motors and gear transmission unit is studied. Then the mathematical model of the cutterhead driving system with consideration of gear frequency cycle error and backlash, and the parameter difference of driving motors is presented. Simulations show that the proposed model provides a useful guidance for improving the control strategy of the cutterhead driving system.

**Keywords:** Cutterhead; Synchronization; Electromechanical coupling.

## 1 Introduction

The shield machine is a kind of technology-intensive major engineering project equipment for tunnel excavation, and its main unit is cutterhead. The rotary driving system of cutterhead generally needs to have the characteristics of high power, high torque output, and continuously bidirectional adjustable speed [1,2]. Because the motors-driving rotary system not only can start at rated load, regulate speed smoothly, but also have high efficiency, and be easy to maintain, it is getting more and more application in shield machines.

The driving system structure is shown in Fig.1. When a mechanical load is driven by multi motors, it requires the motors with consistent dynamic performances and even loads. Nevertheless, the driving system of cutterhead is a strongly coupled nonlinear system. Its gear transmission usually has gear frequency cycle error and gear backlash, which results in limit cycle oscillations, unstable vibration of low speed and turning errors, etc [3,4]. Because of the interaction of gear frequency cycle error, gear backlash and differences in motor parameters, pinion driving shaft load torque distributes unevenly, which also affects the motors' synchronization performance. The sudden load change, manufacturing error or improper control during the

actual project may make one or several small gear shafts suffer sudden increase in torque, even over the ultimate torque, then it will lead to shaft broken. Therefore, the dynamics of the cutterhead driving mechanism and the factors affecting its coordination should be analyzed in depth, and then the methods to improve system coordination should be studied.

Currently, the related researches mainly involve the study of multi-motor synchronized driving servo system, which focus on the impact of the backlash on the system, and the anti-backlash control method [5,6]. However, mechanical and electrical coupling problem is not considered. And with regard to the cutterhead driving system, there is no research on the electromechanical coupling model with consideration of gear frequency cycle error, backlash effect, and differences in motor parameters.

Based on the analysis of the transmission mechanism of the cutterhead driving system, an electromechanical coupling model is build up. And the simulation of a cutterhead driving system is performed with MATLAB/Simulink.

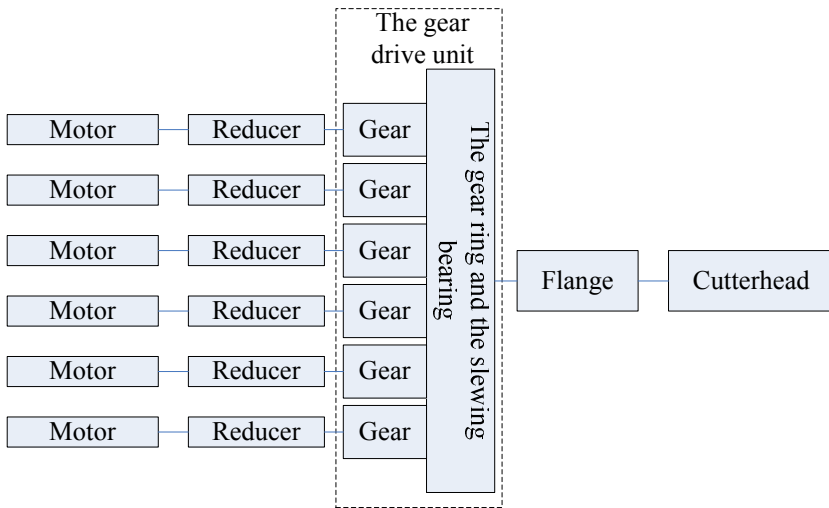


Fig. 1. The structure chart of the cutterhead driving system of shield machines

## 2 Electromechanical Coupling Mathematical Model of the Cutterhead Driving System

Actual operating experience has shown that the main factors affecting the synchronization of the cutterhead driving system are the nonlinearity of transmission mechanism and the driving motors' synchronization performance. In the transmission gear unit, non-linear factors mainly include gear frequency cycle error and gear backlash. On the other hand, parameter differences of the driving motors of the cutterhead have a great impact on synchronization performance. To study multi-motor synchronous control strategy requires the establishment of the system mathematical model taking into account the above factors.

### 2.1 The Kinetic Equation of the Cutterhead Driving System

For the vector controlled motor, the stator current is control variable. Oriented by rotor field in the two-phase synchronous rotating coordinate system (MT coordinate system), the torque formula is

$$T_e = n_p \frac{L_m}{L_r} i_{t1} \psi_2 . \tag{1}$$

where  $T_e$  is electromagnetic torque,  $n_p$  is the number of rotor pole-pairs,  $L_m$  is mutual inductance between stator and rotor,  $L_r$  is rotor self-inductance,  $i_{t1}$  is torque component of stator current,  $\psi_2$  is rotor flux.

For the  $i$ th motor, the motion equation can be expressed as

$$J \dot{\omega}_i = T_{ei} - B \omega_i - \frac{T_i}{h_i} . \tag{2}$$

where  $J$  is moment of inertia,  $\omega$  is angular speed of motor,  $T_{ei}$  is electromagnetic torque corresponding to the  $i$ th motor,  $B$  is friction coefficient,  $T_i$  is the driving torque of the small gear corresponding to the  $i$ th motor, and  $h_i$  is the reduction ratio corresponding to the  $i$ th motor.

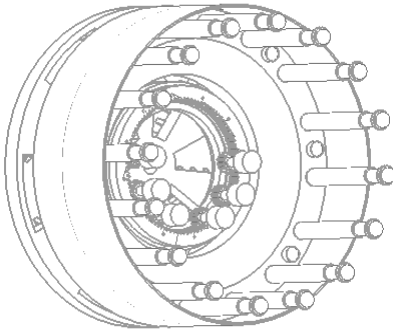


Fig. 2. Shield gear transmission mechanism

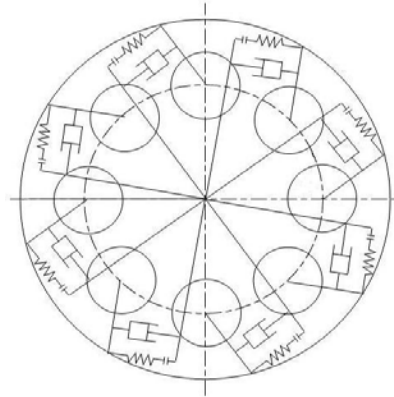


Fig. 3. Gear transmission system model

The transmission mechanism structure of shield machine and its model are shown in Fig.2 and Fig.3 respectively. Due to the manufacturing errors, installation errors, component elastic deformations, lubrication differences, temperature differences and so on, the ideal uniform distribution state of load among the gear train is difficult to achieve. The difference among motors' parameters or motor disturbance will make the uneven load distribution among small gears further intensified.

The dynamic equation of the  $i$ th small gear is

$$\frac{T_i}{r_i} = K_i \cdot f(x_i', b) + m_i \ddot{x}_i + C_i \dot{x}_i . \tag{3}$$

Where,  $i=1, 2, \dots, n$ .  $n$  is the number of small gears, and the  $i$ th small gear corresponding to the  $i$ th motor;  $r_i$  is the radius of small gear;  $K_i$  is the gear meshing coefficient corresponding to the  $i$ th small gear;  $b$  is mesh clearance;  $m_i$  is the quality of the  $i$ th small gear;  $x_i$  is the micro-displacement of the  $i$ th small gear along the acting line;  $C_i$  is the meshing damping coefficient;  $x_i'$  is the relative displacement between the  $i$ th small gear and the ring gear, that is

$$x_i' = x_i - x_{ring} - E_{pri} \sin(\omega_c t + \varphi_{ri}). \tag{4}$$

where  $x_{ring}$  is the micro-displacement of the ring gear along the acting line;  $E_{pri}$  is the equivalent integrated gear frequency error;  $\omega_c$  is the tooth meshing frequency;  $\varphi_{ri}$  is the initial phase. The gear backlash nonlinearity function is

$$f(x_i', b) = \begin{cases} x_i' - b & (x_i' > b) \\ 0 & (-b < x_i' < b) \\ x_i' + b & (x_i' < -b) \end{cases}. \tag{5}$$

Equation (5) can be written as

$$f(x_i', b) = \alpha_i [x_i - x_{ring} - E_{pri} \sin(\omega_c t + \varphi_{ri}) + \beta_i]. \tag{6}$$

where

$$\alpha_i = \begin{cases} 1 & (x_i' > b) \\ 0 & (-b < x_i' < b) \\ 1 & (x_i' < -b) \end{cases}, \beta_i = \begin{cases} -b & (x_i' > b) \\ 0 & (-b < x_i' < b) \\ b & (x_i' < -b) \end{cases}.$$

Therefore, the dynamic equation of the small gear is

$$\frac{1}{r} \begin{bmatrix} T_1 \\ T_2 \\ \dots \\ T_n \end{bmatrix} = \begin{bmatrix} K_1 & 0 & \dots & 0 \\ 0 & K_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & K_n \end{bmatrix} \begin{bmatrix} f(x_1', b) \\ f(x_2', b) \\ \dots \\ f(x_n', b) \end{bmatrix} + \begin{bmatrix} m_1 & 0 & \dots & 0 \\ 0 & m_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & m_n \end{bmatrix} \begin{bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \\ \dots \\ \ddot{x}_n \end{bmatrix} + \begin{bmatrix} C_1 & 0 & \dots & 0 \\ 0 & C_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & C_n \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dots \\ \dot{x}_n \end{bmatrix} \tag{7}$$

And the dynamic equation of the ring gear is

$$\frac{T_{load}}{r_{ring}} = m_{ring} \ddot{x}_{ring} - \sum_{i=1}^n K_i \cdot f(x_i', b). \tag{8}$$

where  $T_{load}$  is the load torque;  $r_{ring}$  is the radius of ring gear;  $m_{ring}$  is the mass of ring gear.

In summary, the state equation of the cutterhead driving system with consideration of gear frequency cycle error and backlash effect is

$$\begin{aligned}
 & \begin{bmatrix} \dot{\omega} \\ \dot{x}_1 \\ \dot{\omega}_2 \\ \dot{x}_2 \\ \vdots \\ \dot{\omega}_n \\ \dot{x}_n \\ \dot{\omega}_{ring} \\ \dot{x}_{ring} \end{bmatrix} = \begin{bmatrix} \frac{Bh_1+C_1r_1^2}{h_1J_1+J_{c1}} & \frac{K_1r_1\alpha_1}{h_1J_1+J_{c1}} & 0 & 0 & \dots & \dots & 0 & 0 & 0 & \frac{K_1r_1\alpha_1}{h_1J_1+J_{c1}} \\ r_1 & 0 & 0 & 0 & \dots & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{Bh_2+C_2r_2^2}{h_2J_2+J_{c2}} & \frac{K_2r_2\alpha_2}{h_2J_2+J_{c2}} & \dots & \dots & 0 & 0 & 0 & \frac{K_2r_2\alpha_2}{h_2J_2+J_{c2}} \\ 0 & 0 & r_2 & 0 & \dots & \dots & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \dots & \frac{Bh_n+C_nr_n^2}{h_nJ_n+J_{cn}} & \frac{K_nr_n\alpha_n}{h_nJ_n+J_{cn}} & 0 & \frac{K_nr_n\alpha_n}{h_nJ_n+J_{cn}} \\ 0 & 0 & 0 & 0 & \dots & \dots & r_n & 0 & 0 & 0 \\ 0 & \frac{r_{ring}K_1\alpha_1}{J_{ring}} & 0 & \frac{r_{ring}K_2\alpha_2}{J_{ring}} & \dots & \dots & 0 & \frac{r_{ring}K_n\alpha_n}{J_{ring}} & \frac{B_{ring}}{J_{ring}} & \sum_{i=1}^n \frac{r_{ring}K_i\alpha_i}{J_{ring}} \\ 0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 & r_{ring} & 0 \end{bmatrix} \begin{bmatrix} \omega \\ x_1 \\ \omega_2 \\ x_2 \\ \vdots \\ \omega_n \\ x_n \\ \omega_{ring} \\ x_{ring} \end{bmatrix} \\
 & + \begin{bmatrix} \frac{K_1r_1\alpha_1}{h_1J_1+J_{c1}} & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & \frac{K_2r_2\alpha_2}{h_2J_2+J_{c2}} & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{K_nr_n\alpha_n}{h_nJ_n+J_{cn}} & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} \begin{bmatrix} E_{p1}\sin(\omega t+\varphi_1)-\beta_1 \\ E_{p2}\sin(\omega t+\varphi_2)-\beta_2 \\ \vdots \\ E_{pn}\sin(\omega t+\varphi_n)-\beta_n \\ \frac{r_{ring}}{J_{ring}} \sum_{i=1}^n K_i\alpha_i[\beta_i-E_{pi}\sin(\omega t+\varphi_i)] \end{bmatrix} \quad (9) \\
 & + \begin{bmatrix} h_{n1} \frac{L_{n1}}{L_{r1}} i_{11} \psi_{2,1} & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & h_{n2} \frac{L_{n2}}{L_{r2}} i_{12} \psi_{2,2} & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & h_{nm} \frac{L_{nm}}{L_m} i_{1m} \psi_{2,m} & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} \begin{bmatrix} i_{n,1} \\ i_{n,2} \\ \vdots \\ i_{n,n} \\ -T_{load} \end{bmatrix}
 \end{aligned}$$

The model takes into account the difference of parameters among the motors. Meshing movement is achieved by the interaction of gear elastic and viscous friction. But in general, viscous friction is negligible. In addition, due to manufacturing errors and mechanical errors will cause gear mesh clearance, therefore the frequency error and tooth backlash effect is considered.

### 2.2 Multi-motor Synchronous Control Strategy

Currently, most cutterhead driving system of the shield machine uses master-slave control strategy combined with current balanced control method [7,8]. The control principle is to retain the current loop of every driving motor, make the current loop signals of the master and the slaves in parallel, and make the motors' dynamic response the same by adjusting the corresponding parameters of the motors' current loop, so that the slaves' torque can follow the master's torque and ensure multi-motor operation synchronized and the load distribution balanced automatically. The system structure is shown in Fig. 4.

As shown in Fig.4, ASR is the speed controller, and ATR is the torque controller (generally using PI in actual projects). The motor #1 works as the master motor, and the others is the slaves. The speed loop of main motor retains complete, and the slaves work at the speed open-loop state. All the slave motors achieve speed synchronization with the master through the gears meshing.

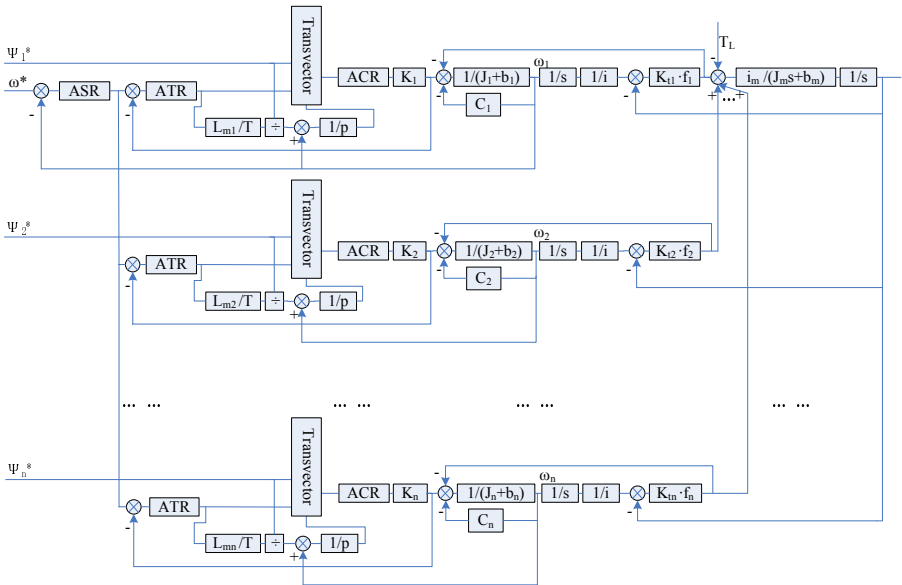
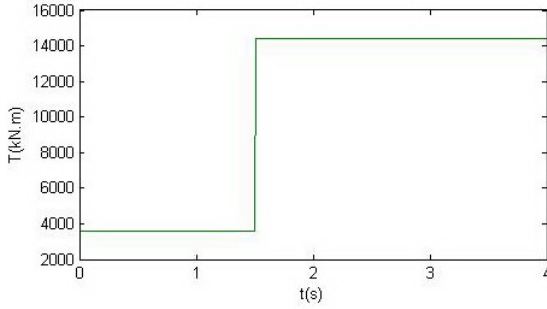


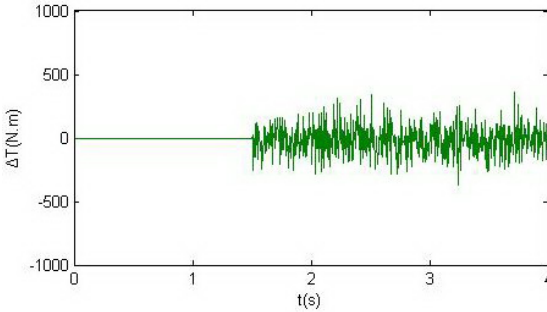
Fig. 4. The control structure of the cutterhead driving system using master-slave control strategy

### 3 Simulation Analysis

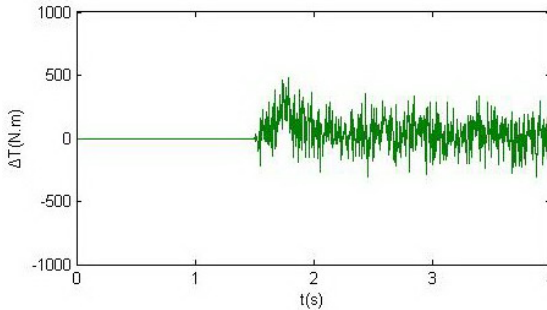
The electromechanical coupling simulation model of the cutterhead driving system is build under Matlab. The shield machine for analysis includes six driving motors with the master-slave control strategy. Let the motor #1 is the master. There exist differences among the motors' parameters that the master's stator resistance value is 3% higher than others. The rated power of the motor is 190kW, and the reducer reduction ratio is 1:100. The diameter of the shield cutterhead is 11.97m, the rated torque is 14000 kN·m, and the number of gear ring and gear teeth is 174 and 14 respectively.



**Fig. 5.** The snap load of the cutterhead which changes from 3600 kN·m to 14500 kN·m at the 3rd second



**Fig. 6.** The torque difference between the master motor and the slaves when the comprehensive gear frequency error amplitude is 100μm



**Fig. 7.** The torque difference between the master motor and the slaves when the comprehensive gear frequency error amplitude is 300μm

The cutterhead load has the step change(as shown in Fig.5). The backlash value is 100μm. And the comprehensive gear frequency error amplitude is 100μm and 300μm respectively, the corresponding torque differences between the master motor and the slaves are shown in Fig.6 and Fig.7 respectively. As shown in the figures, with the



increase of gear frequency error magnitude, the balance effect of motors' load come worse. The larger the gear frequency error is, the greater the torque difference become when the cutterhead load change suddenly. The results agree with the actual situation.

Using the master-slave control strategy, the gear frequency cycle error has a great impact on the balance the transmission drive shaft torque. The instantaneous mutation of the load on the cutterhead because of qualitative changes (as shown in Fig.5) makes that the drive shaft corresponding to the main motor gets the most instantaneous sharing of the whole load torque (as shown in Fig.6 and Fig.7). And due to the torque master-slave control strategy, the drive shaft torque corresponding to the slave motors rises instantaneously in pace with the master motor. When the instantaneous torque exceeds the allowable torque of the drive shafts, series of off-axis phenomenon may occur.

## 4 Conclusion

The cutterhead driving system is a multivariable, strong coupled, complicated nonlinear system. The existing of gear frequency error, backlash and motors' parameters difference affects the system coordination. The paper studies systematically the electromechanical coupling relationship of the system, and establishes the mathematical model of the cutterhead driving system with consideration of the influencing factors above, which is simulated under Matlab platform. Simulation results show that the model can reflect the impact of gear frequency error, backlash and motors' parameters difference on the system coordination. It lays the foundation for the further research on the simulation, analysis approach to improving the synchronization coordination of the cutterhead driving system.

## References

1. Shi, Y., Zhuang, Q., Lv, J.: Design of Integral Driving System for Cutter Head of Shield Tunneling Machine. *Road Machinery & Construction Mechanization* 25, 18–20 (2008)
2. Wang, J.Y.: Tunneling and Technological Progress in Tunneling in China. *Tunnels and Underground Structures* 5, 97–109 (2000)
3. Li, R.-F., Wang, J.-J.: *Gear System Dynamics*. Science Press, Beijing (1997)
4. Cheng, Y.P., Lim, T.C.: Dynamics of hypoid gear transmission with nonlinear time-varying mesh characteristics. *Journal of Mechanical Design, Transactions of the ASME*, 373–382 (2003)
5. Orlov, Y., Aguilar, L., Cadiou, J.C.: Switching chattering control vs. backlash/friction phenomena in electrical servo-motors. *Int. Journal of Control*. 76, 959–967 (2003)
6. Jukic, T., Peric, N.: Model based backlash compensation. In: *Proc. of 2001 American Control Conf.*, Arlington, VA, USA, pp. 775–780 (2001)
7. Tang, J., Li, Z.-Y.: Application of Variable Frequency Speed Regulation in Double Motor Synchronization Drive. *Electric Drive of China* 11, 8–11 (2008)
8. Chen, Q.-W., Guo, Y., Yang, F., Hu, W.-L.: Double-motor Synchronizing Drive Control System. *Journal of Nanjing University of Science and Technology* 29, 103–107 (2005)

# Dimension Optimization of an Orientation Fine-Tuning Manipulator for Segment Assembly Device

Chao Wu, Xin-Jun Liu\*, and Jinsong Wang

Department of Precision Instruments and Mechanology, Tsinghua University,  
Beijing, 100084 China  
wu-c06@mails.tsinghua.edu.cn,  
{xinjunliu,wangjs}@mail.tsinghua.edu.cn

**Abstract.** This study focuses on the dimension synthesis of a three-DOF spherical parallel manipulator which can be used for orientation fine-tuning of segments in shield tunneling machine. Several performance indices are introduced by taking into account the motion/force transmissibility. The workspace of the manipulator is discussed. The process of determining the optimum geometric parameters is presented with respect to the performance chart. The manipulator that has good motion/force transmissibility in the whole workspace is identified.

**Keywords:** spherical parallel manipulator, transmission index, segment assembly.

## 1 Introduction

In the field of development of underground space, shield tunneling machines have been widely used in subway tunnels, channel tunnels, and other kinds of tunnels. Segment assembly device is one of the key sub-devices of shield tunneling machine, and its effectiveness has a large influence on the rate of advance. Since the segment has a large weight, it is hard to precisely achieve its required pose. This will seriously lower the efficiency of segment assembly. Thus, orientation fine-tuning of the segments is necessary for shield tunneling machine. As is well known, parallel manipulators have the advantages of a compact structure, high stiffness, and a high load/weight ratio. A spherical three-degree-of-freedom (DOF) parallel manipulator can be used as the orientation fine-tuning manipulator for segment assembly device in shield tunneling machine to adjust the orientation of a segment about any axis.

For spherical three-DOF parallel manipulators, there are many types of structures. One of the most popular structures is the 3-SPS-1-S manipulator as shown in Fig. 1(a). Here, S and P stand for spherical and prismatic joints, respectively, and the joint with an underlined symbol is actuated. Many efforts have been contributed to this manipulator. Ji and Wu [1] studied the forward kinematics of the 3-DOF spherical parallel manipulator. Several works [2], [3], [4], etc. investigated the singularity of the manipulator, and presented the computation and representation of the singularity loci. In [5], topology optimization of the 3-SPS-1-S manipulator was studied to determine design parameters

---

\* Corresponding author.

based on the condition number of the manipulator Jacobian matrix over the entire workspace. In the design of parallel manipulators, the most commonly used performance indices are the *local conditioning index* (LCI) [6] and the *global conditioning index* (GCI) [7]. However, a recent study [8] found serious inconsistencies when these indices were applied to the design of mixed-DOF parallel manipulators (those employing both translations and rotations). Wang et al. [9] found that the LCI was ineffective when applied to a planar parallel manipulator with only translational DOFs. To eliminate singularity and its near configurations, the LCI was used by most researchers. A good-condition [10] or effective workspace [11] is usually defined in terms of a specified minimum LCI. This minimum LCI is arbitrary since a definite value cannot be assigned. It is not possible to define a mathematical distance from the current pose to a singular configuration for a parallel manipulator by using LCI. The authors therefore think that it should be careful to use LCI in the design of parallel manipulators. Being counterparts of serial robots, parallel manipulators are always good at motion/force transmission but not just dexterous manipulation. An index that evaluates the motion/force transmissibility of parallel manipulators should be figured out.

In this paper, several transmission indices will be introduced to evaluate the motion/force transmissibility of the 3-SPS-1-S parallel manipulator. Dimensional optimization of the manipulator will be achieved by using these indices. During the dimension optimization, the parameter design space is first established, performance chart is then plotted. The process of determining the optimum geometric parameters is presented with respect to the performance chart.

## 2 Description of the Orientation Fine-Tuning Manipulator

### 2.1 Architecture

Figure 1(a) shows an orientation fine-tuning manipulator. Since its moving platform is connected to the fixed platform through three identical SPS legs and one passive S joint, it can be referred to as the 3-SPS-1-S parallel manipulator. The three P joints are actuated. The moving platform can only rotate along an axis passing through the center of the passive S joint, thus the center is referred to as the rotation center of the moving platform.

As shown in Fig. 1(b),  $A_i$  and  $B_i$  ( $i=1, 2, 3$ ) are the centers of the S joints that connected to the moving and fixed platforms.  $A_1A_2A_3$  and  $B_1B_2B_3$  are both two equilateral triangles. The origins of the global and moving coordinate systems, i.e.,  $O-xyz$  and  $O-x'y'z'$ , are both at the rotation center  $O$ . The moving coordinate system  $O-x'y'z'$  is attached to the moving platform. The  $z'$ -axis passes through the center of the moving platform and is perpendicular to the equilateral triangle  $B_1B_2B_3$ . The S joint centers  $A_i$  and  $B_i$  always lie in the planes  $O-xz$  and  $O-x'z'$ , respectively. There are four geometric parameters in the manipulator. Since the fixed and moving platforms are both equilateral triangles, the radiuses of their circumcircles are expressed by  $r_1$  and  $r_2$ , respectively. Parameters  $h_1$  and  $h_2$  represent the distances from the origin  $O$  to the planes  $A_1A_2A_3$  and  $B_1B_2B_3$ , respectively.

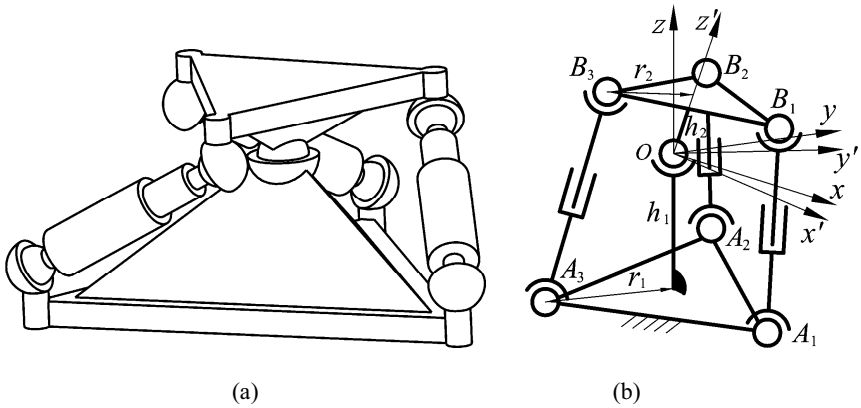


Fig. 1. An orientation fine-tuning manipulator: (a) kinematic structure; (b) kinematic scheme.

### 2.2 Orientation Representation

During the orientation fine-tuning of segments, the manipulator usually needs to achieve some required motions, i.e., roll, pitch, and yaw. Hereby, the Roll-Pitch-Yaw angles  $(\alpha, \beta, \gamma)$  are used to represent the orientation of the moving platform. With respect to the global coordinate system  $O - xyz$ , the rotation matrix can be expressed by

$$\begin{aligned}
 \mathbf{R} &= \mathbf{R}(Z, \alpha)\mathbf{R}(Y, \beta)\mathbf{R}(X, \gamma) \\
 &= \begin{bmatrix} c\alpha c\beta & c\alpha s\beta s\gamma - s\alpha c\gamma & c\alpha s\beta c\gamma + s\alpha s\gamma \\ s\alpha c\beta & s\alpha s\beta s\gamma + c\alpha c\gamma & s\alpha s\beta c\gamma - c\alpha s\gamma \\ -s\beta & c\beta s\gamma & c\beta c\gamma \end{bmatrix} \tag{1}
 \end{aligned}$$

where  $c$  and  $s$  represent the operators of cosine and sine, respectively.

## 3 Definition of Transmission Indices

The function of a mechanism is to transmit motion/force from its input member to its output member. Mechanisms should transmit force in order to balance the payload on the output member. During force transmission, the arising internal wrench, namely the transmission wrench, can be expressed by the *transmission wrench screw* (TWS).

Transmission indices must be defined in order to evaluate the motion/force transmission performance of the manipulator. Screw theory will be employed in this study to define the motion/force transmission indices. Then, the screw theory will be briefly introduced in the next section.

### 3.1 Foundation of Screw Theory

The instantaneous motion of a rigid body can be represented by using a twist screw. A twist screw can be expressed in the form of Plücker coordinates as

$$\mathcal{S}_1 = w\mathcal{S}_1 = w(\boldsymbol{\omega}_1; \boldsymbol{\nu}_1) = (L_1, M_1, N_1; P_1^*, Q_1^*, R_1^*), \quad (2)$$

where  $w$  is the amplitude of the twist screw  $\mathcal{S}_1$ , and  $\mathcal{S}_1$  is the unit twist screw or the normalized twist screw ( $\mathcal{S}$  in italics represents a unit screw). A wrench exerted on the rigid body can be expressed as a wrench screw in the form of Plücker coordinates as

$$\mathcal{S}_2 = f\mathcal{S}_2 = f(\boldsymbol{f}_2; \boldsymbol{\tau}_2) = (L_2, M_2, N_2; P_2^*, Q_2^*, R_2^*), \quad (3)$$

where  $f$  is the amplitude of the wrench screw  $\mathcal{S}_2$  and  $\mathcal{S}_2$  is the unit wrench screw or the normalized wrench screw.

The reciprocal product between the two unit screws  $\mathcal{S}_1$  and  $\mathcal{S}_2$  is

$$\mathcal{S}_1 \circ \mathcal{S}_2 = \boldsymbol{f}_2 \cdot \boldsymbol{\nu}_1 + \boldsymbol{\tau}_2 \cdot \boldsymbol{\omega}_1 = L_1 P_2^* + M_1 Q_2^* + N_1 R_2^* + L_2 P_1^* + M_2 Q_1^* + N_2 R_1^*. \quad (4)$$

This defines the instantaneous power between the unit wrench and the unit twist. When the reciprocal product of  $\mathcal{S}_1$  and  $\mathcal{S}_2$  equals zero, the unit wrench represented by  $\mathcal{S}_2$  will apply no work to the rigid body. The rigid body therefore moves along the unit twist screw  $\mathcal{S}_1$ .

To evaluate the effectiveness of motion/force transmission, an index is defined as

$$\rho = \frac{|\mathcal{S}_1 \circ \mathcal{S}_2|}{|\mathcal{S}_1 \circ \mathcal{S}_2|_{\max}}. \quad (5)$$

Here  $|\mathcal{S}_1 \circ \mathcal{S}_2|_{\max}$  represents the potential maximal magnitude of the reciprocal product of  $\mathcal{S}_1$  and  $\mathcal{S}_2$ . The larger the value of  $\rho$ , the more effectively the power is transmitted from the wrench to the twist. This indicates a better motion/force transmission. Hereby,  $\rho$  is referred to as the *power coefficient*.

### 3.2 Local Transmission Index (LTI)

For the 3-SPS-1-S manipulator, there is a transmission wrench in each SPS leg. In leg  $i$ , the transmission wrench is a pure force along the line  $\overline{A_i B_i}$ , and the unit TWS can be expressed by

$$\mathcal{S}_{Ti} = (\boldsymbol{t}_i; \boldsymbol{a}_i \times \boldsymbol{t}_i) = (\boldsymbol{t}_i; \boldsymbol{b}_i \times \boldsymbol{t}_i) \quad (i=1,2,3), \quad (6)$$

where  $\boldsymbol{t}_i$  stands for the unit vector along the line  $\overline{A_i B_i}$ , and  $\boldsymbol{a}_i$  and  $\boldsymbol{b}_i$  represent the vectors from origin  $O$  to the S joint centers  $A_i$  and  $B_i$ , respectively.

When keeping the P joint in leg  $i$  actuated and fixing the other two P joints, only the power from the P joint in leg  $i$  can be transmitted to the moving platform, and the transmission wrenches in other two legs become constraint wrenches. The manipulator can then be considered as a single-DOF mechanism. In this case, the instantaneous unit motion of the moving platform is represented by a unit twist screw  $\mathcal{S}_{O_i}$ .

The *power coefficient* between the TWS  $\$_{Ti}$  and the twist screw  $\$_{Oi}$  is defined as the transmission index of leg  $i$ . It can be represented by

$$\eta_i = \frac{|\$_{Ti} \circ \$_{Oi}|}{|\$_{Ti} \circ \$_{Oi}|_{\max}} \quad (i = 1, 2, 3). \tag{7}$$

Here, the determination of the index  $\eta_i$  is presented. When fixing the P joints in the second and third legs, the two transmission wrenches in these legs become constraint wrenches for the moving platform. In other words, two more constraint wrenches are exerted on the moving platform. Actually, the two constraint wrenches can be considered as two constraint moments, and the constraint wrench screws can be represented by

$$\$_{T2}^r = (0; \mathbf{a}_2 \times \mathbf{b}_2), \tag{8}$$

$$\$_{T3}^r = (0; \mathbf{a}_3 \times \mathbf{b}_3). \tag{9}$$

Then the instantaneous unit motion of the moving platform can be represented by

$$\$_{O1} = ((\mathbf{a}_2 \times \mathbf{b}_2) \times (\mathbf{a}_3 \times \mathbf{b}_3) / |(\mathbf{a}_2 \times \mathbf{b}_2) \times (\mathbf{a}_3 \times \mathbf{b}_3)|; 0). \tag{10}$$

We may get

$$|\$_{T1} \circ \$_{O1}| = \frac{|(\mathbf{a}_2 \times \mathbf{b}_2) \times (\mathbf{a}_3 \times \mathbf{b}_3) \cdot (\mathbf{a}_1 \times \mathbf{t}_1)|}{|(\mathbf{a}_2 \times \mathbf{b}_2) \times (\mathbf{a}_3 \times \mathbf{b}_3)|}. \tag{11}$$

Since the axis of  $\$_{T1}$  always passes through the S joint center  $B_1$ , the potential maximal magnitude of the reciprocal product of  $\$_{T1}$  and  $\$_{O1}$  is equal to the distance from  $B_1$  to the axis of  $\$_{O1}$ . As shown in Fig. 2,  $d_{\max}$  represents the distance from  $B_1$  to the axis of  $\$_{O1}$ . Then we get

$$|\$_{T1} \circ \$_{O1}|_{\max} = d_{\max}. \tag{12}$$

Substituting (11) and (12) to (7), the index  $\eta_1$  can be obtained. The other two transmission indices can be determined in the same way.

When one of the transmission indices is small, the manipulator has poor motion/force transmissibility. When all of the input and output transmission indices are large, the manipulator is said to be good at motion/force transmission. The transmission index for the whole manipulator is then defined as

$$\lambda = \min\{\eta_i\}, \quad (i = 1, 2, 3). \tag{13}$$

Since  $\lambda$  will be different in different configurations,  $\lambda$  is referred to as the *local transmission index* (LTI). The value of  $\lambda$  ranges from 0 to 1. A larger  $\lambda$  indicates better motion/force transmission.

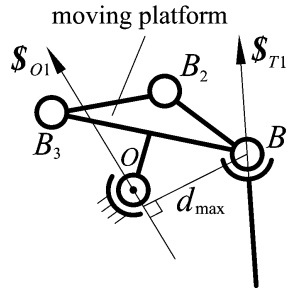


Fig. 2. Relationship between TWS and output twist screw

### 3.3 Global Transmission Index (GTI)

The LTI can only judge the effectiveness of motion/force transmission in a single pose. A parallel manipulator is usually designed to achieve a set of poses instead of one. Therefore, the behavior of the manipulator in a set of poses should be considered. Then, the *global transmission index* (GTI) is defined based on the definition of GCI in [11] as

$$\Gamma = \frac{\int_W \lambda dW}{\int_W dW}, \quad (14)$$

where  $W$  represents the workspace of the moving platform (it will be discussed in the next section). A larger  $\Gamma$  usually indicates better motion/force transmissibility during the whole workspace.

## 4 Workspace

In the definition of workspace, the first problem is how to determine the original orientation of the moving platform. For the 3-SPS-1-S manipulator, the moving platform is usually parallel to the fixed platform at its original orientation. In this study, the moving platform is defined at its original orientation when the pitch and yaw angles are both zero, i.e.,  $\beta = \gamma = 0^\circ$ . Then the roll angle  $\alpha$  needs to be determined.

By giving the manipulator a group of parameters as  $r_1 = 1.2$ ,  $r_2 = 0.8$ ,  $h_1 = 1.0$ , and  $h_2 = 0$ , Fig. (3) illustrates the relationship between LTI and roll angle  $\alpha$ . It can be seen from Fig. (3) that the LTI reaches the maximum 0.9925 at  $\alpha = 39^\circ$ . It means that  $(39^\circ, 0^\circ, 0^\circ)$  is the best original orientation for this manipulator. With different geometric parameters, however, the best value of  $\alpha$  for the original orientation will be different. Hereby, when the geometric parameters of the manipulator are given, the roll angle of the original orientation is defined as the angle at which the LTI has the maximum value by fixing  $\beta = \gamma = 0^\circ$ .

Since the 3-SPS-1-S manipulator is used for orientation fine-tuning of the segment, there is no need to have a very large workspace. Hereby, the workspace of the manipulator is defined as  $\alpha \in [\alpha_0 - 20^\circ, \alpha_0 + 20^\circ]$ ,  $\beta \in [-20^\circ, 20^\circ]$ , and  $\gamma \in [-20^\circ, 20^\circ]$ , where  $\alpha_0$  represents the roll angle of the original orientation.

### 5 Dimension Optimization

It is well known that the performance of a mechanism is closely related to its geometric parameters. In this section, the process of optimizing the geometric parameters is presented to enable the manipulator to achieve good motion/force transmission performance during the workspace.

For the 3-SPS-1-S parallel manipulator, there are four geometric parameters,  $r_1$ ,  $r_2$ ,  $h_1$  and  $h_2$ . In this study, a class of 3-SPS-1-S manipulators with  $h_2 = 0$  is only investigated since the manipulator with  $h_2 = 0$  will have a relative small space volume. Then there are three parameters, i.e.,  $r_1$ ,  $r_2$ , and  $h_1$  left to be optimized.

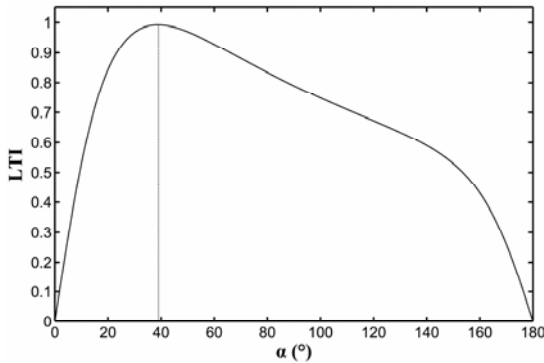


Fig. 3. Relationship between LTI and the roll angle  $\alpha$

#### 5.1 Parameter Design Space

In order to show the relationship between performance and the design parameters, the design concept introduced in [12] is extended to the 3-SPS-1-S manipulator. The first step is to set up the parameter design space.

As stated in [12], the performance of a manipulator with  $r_1$ ,  $r_2$ , and  $h_1$  may be similar to or identical with that of a manipulator with  $Dr_1$ ,  $Dr_2$ , and  $Dh_1$ , where  $D$  is any positive number. For the manipulator studied here, all the three parameters  $r_1$ ,  $r_2$ , and  $h_1$  can be considered as non-dimensional parameters. Letting them meet the following equation [12], i.e.,

$$r_1 + r_2 + h_1 = 3. \tag{15}$$



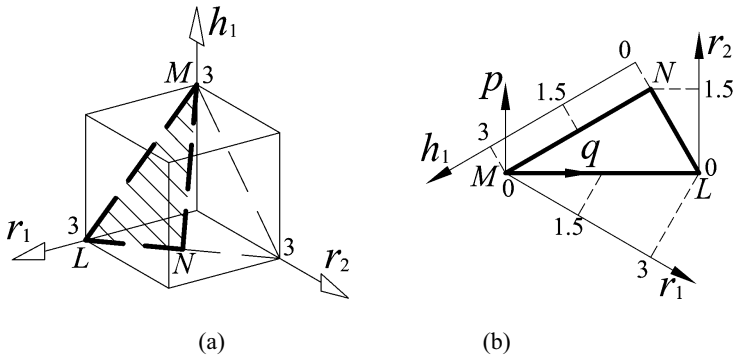


Fig. 4. Parameter design space of the 3-SPS-1-S parallel manipulator

Then it is possible to reduce a three-dimensional problem to a two-dimensional one. For the manipulator studied here, parameters  $r_1$ ,  $r_2$ , and  $h_1$  should be

$$\begin{cases} 0 < r_1, r_2, h_1 < 3 \\ r_2 < r_1 \end{cases} \quad (16)$$

Based on (15) and (16), the parameter design space can be established as the triangle  $LMN$  (shown in Fig. 4(a)). Meanwhile, it can be changed to a two-dimensional form, which makes it easier to plot an atlas. By using

$$\begin{cases} p = r_2 \\ q = \sqrt{3} + r_1/\sqrt{3} - h_1/\sqrt{3} \end{cases} \quad (17)$$

two orthogonal coordinates  $p$  and  $q$  can be utilized to express  $r_1$ ,  $r_2$ , and  $h_1$ . The design space in planar configuration is the triangle  $LMN$  as shown in Fig. 4(b).

### 5.2 Atlas of GTI

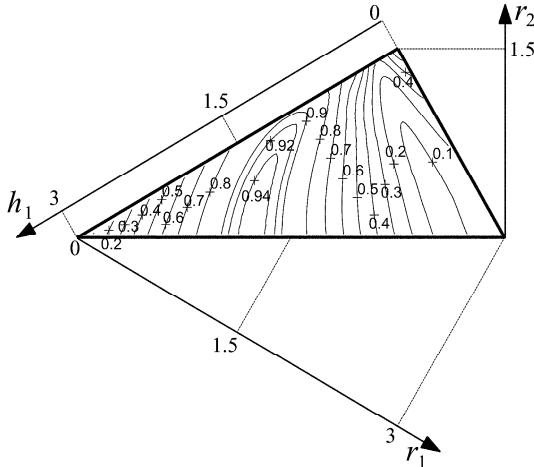
Based on the definition of GTI in Section 3.3, it is possible to obtain the GTI of any non-dimensional manipulator in the parameter design space. The atlas shown in Fig. 5 can then be plotted to illustrate the relationship between GTI and geometric parameters  $r_1$ ,  $r_2$ , and  $h_1$ .

### 5.3 Optimization Process

In this section, the optimization process based on the atlas in Fig. 5 can be summarized as below:

**Step 1:** Identify an optimum region in the parameter design space. Letting  $GTI \geq 0.94$ , an optimum region can be identified (shown as the hatched region in Fig. 6) by using the atlas in Fig. 5. The optimum region includes all possible solutions of the non-dimensional parameters  $r_1$ ,  $r_2$ , and  $h_1$ . It should be noted that the optimum region is dependent upon the design requirement. In other words, the optimum region can be changed when the design requirement is different.

**Step 2:** Select a solution from the optimum region. Although all the solutions in the optimum region meet the design requirement  $GTI \geq 0.94$ , one of them should be selected for the geometric parameters of the manipulator. Actually besides  $GTI$ , practical situation of the application should always be taken into account for the selection of the solution.



**Fig. 5.** Atlas of  $GTI$  of the 3-SPS-1-S parallel manipulator

As mentioned previously, the 3-SPS-1-S parallel manipulator is used for the orientation adjustment of the segment. Thus the manipulator is part of the segment assembly machine. Considering the limited space in the shield tunneling system, the volume of the segment assembly machine should not be very large, then the occupation space of the 3-SPS-1-S manipulator should be as small as possible. This means parameters  $r_1$  and  $h_1$  should be comparatively smaller. Meanwhile, a holding mechanism for segments will be fixed in the moving platform, thus the area of the moving platform should be large enough. In other words,  $r_2$  should not be very small. For these reasons, the manipulator with parameters  $r_1 = 1.05$ ,  $r_2 = 0.6$ , and  $h_1 = 1.35$  is selected from the optimum region here. Then we may get  $r_1 : r_2 : h_1 = 7 : 4 : 9$ .

**Step 3:** Determine the dimensional parameters  $R_1$ ,  $R_2$  and  $H_1$ . In this step, a factor  $D$  is needed to be determined first [12]. This factor can be obtained by comparing the practical space provided for the manipulator to the occupation space of the non-dimensional manipulator selected from the optimum region. For the 3-SPS-1-S manipulator studied here, its occupation space is mainly dependent on the radius ( $R_1$ ) of the circumcircle of fixed platform. Hereby, we suppose that  $R_1 = 700\text{mm}$ . Then, we get  $D = R_1/r_1 = 666.67\text{mm}$ . Accordingly,  $R_2 = D \cdot r_2 = 400\text{mm}$  and  $H_1 = D \cdot h_1 = 900\text{mm}$ .

**Step 4:** Check whether the dimensions obtained in Step 3 is fit for a practical design or not. In Step 3, the dimensional parameters were determined as  $R_1 = 700\text{mm}$ ,

$R_2 = 400\text{mm}$  and  $H_1 = 900\text{mm}$ . In this step, it is necessary to check the dimensions to make sure that the holding mechanism can be fixed in the moving platform well and any other item needed in practical design. If it works, proceed to Step 5; otherwise, return to Step 2, and select another group of non-dimensional parameters from the optimum region, then repeat Steps 3 and 4.

**Step 5:** Calculations of the corresponding index and input range. For the 3-SPS-1-S manipulator with  $R_1 = 700\text{mm}$ ,  $R_2 = 400\text{mm}$ ,  $H_1 = 900\text{mm}$  and  $H_2 = 0\text{mm}$ , the original orientation of the moving platform can be obtained as  $(60^\circ, 0^\circ, 0^\circ)$  based on the definition in Sec. 4. When the moving platform is at its original orientation, the LTI of the manipulator reaches its maximal value 0.9979. The GTI of the manipulator is 0.9428. As to the input range, the link lengths  $|A_i B_i| (i=1,2,3)$  yield to  $[1015.4\text{mm}, 1167.4\text{mm}]$ .

Notably, if the input range is not suitable for a commercial actuator, return to Step 2 and select another group of non-dimensional parameters, or return to Step 1 and identify another optimum region.

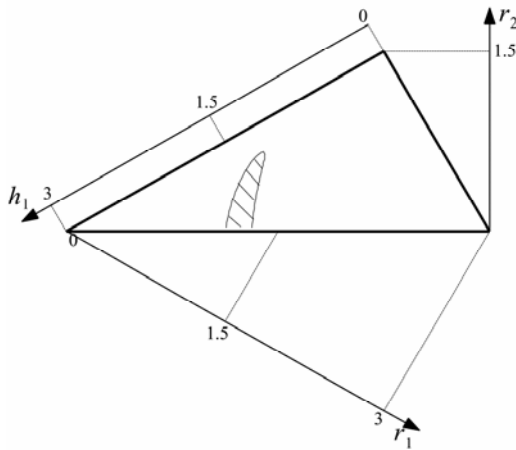


Fig. 6. An optimum region for the 3-SPS-1-S parallel manipulator when  $\text{GTI} \geq 0.94$

## 6 Conclusion

This paper addressed the dimension optimization of the 3-SPS-1-S parallel manipulator which can be used as an orientation fine-tuning manipulator for segment assembly device in shield tunneling machine. Two performance indices, i.e., *local transmission index* (LTI) and *global transmission index* (GTI), were proposed by taking into account the motion/force transmissibility of the manipulator. The original orientation of the moving platform was defined, based on which the workspace was determined. The GTI atlas in parameter design space was illustrated. By giving a design requirement of GTI, an optimum region including all possible optimum solutions for a requirement can be identified from the design space. In addition, considering practical

situation of the application, a group of geometric parameters were selected from the optimum region.

**Acknowledgments.** This work was supported in part by the National Basic Research Program (973 Program) of China (No. 2007CB714000), the High Technology Research and Development Program (863 Program) of China (No. 2010AA101404), and by Program for New Century Excellent Talents in University of China (No. NCET-08-0323).

## References

1. Ji, P., Wu, H.: Algebraic solution to forward kinematics of a 3-DOF spherical parallel manipulator. *Journal of Robotic Systems* 18, 251–257 (2001)
2. Wang, J., Gosselin, C.M.: Singularity loci of a special class of spherical 3-DOF parallel mechanisms with prismatic actuators. *Journal of Mechanical Design* 126, 319–325 (2004)
3. Bonev, I.A., Gosselin, C.M.: Singularity loci of spherical parallel mechanisms. In: *Proceedings of IEEE International Conference on Robotics and Automation, Barcelona, Spain*, pp. 2957–2962 (2005)
4. Bonev, I.A., Gosselin, C.M.: Analytical determination of the workspace of symmetrical spherical parallel mechanisms. *IEEE Transaction on Robotics* 22, 1011–1017 (2006)
5. Alici, G., Shirinzadeh, B.: Topology optimisation and singularity analysis of a 3-SPS parallel manipulator with a passive constraining spherical joint. *Mechanism and Machine Theory* 39, 215–235 (2004)
6. Gosselin, C.M., Angeles, J.: The optimum kinematic design of a spherical three-degree-of-freedom parallel manipulator. *J. Mech. Transm. Autom. Des.* 111, 202–207 (1989)
7. Gosselin, C.M., Angeles, J.: A global performance index for the kinematic optimization of robotic manipulators. *Journal of Mechanical Design* 113, 220–226 (1991)
8. Merlet, J.P.: Jacobian, manipulability, condition number, and accuracy of parallel robots. *Journal of Mechanical Design* 128, 199–206 (2006)
9. Wang, J., Liu, X.-J., Wu, C.: Optimal design of a new spatial 3-DOF parallel robot with respect to a frame-free index. *Science in China, Series E-Technological Sciences* 52, 986–999 (2009)
10. Liu, X.-J., Guan, L., Wang, J.: Kinematics and Closed Optimal Design of a Kind of PRRRP Parallel Manipulator. *Journal of Mechanical Design* 129, 558–563 (2007)
11. Lou, Y.J., Liu, G.F., Chen, N., Li, Z.X.: Optimal design of parallel manipulators for maximum effective regular workspace. In: *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Canada*, pp. 1208–1213 (2005)
12. Liu, X.-J., Wang, J.: A new methodology for optimal kinematic design of parallel mechanisms. *Mechanism and Machine Theory* 42, 1210–1224 (2007)

# Rotation and Gate Movement Compound Control of Screw Conveyor Hydraulic System with Disturbance Compensation

Lintao Wang, Guofang Gong, Hu Shi, and Huaiyin Liu

Zhejiang University, State Key Laboratory of Fluid Power Transmission and Control,  
Hangzhou, Zheda Road No. 38, 310027, China  
jixiezhixing2005@163.com

**Abstract.** This paper presents a new electro-hydraulic control system for screw conveyor device of shield tunnelling machine. The control model of the system is established, in which rotation and gate movement compound control approach is applied. It utilizes the motor speed feedback to design an outer loop and cylinder velocity feedback as an inner loop to realize flow rate disturbance compensation. Simulations are done and results are presented to verify the effectiveness and rationality of the proposed driving system and its control strategy.

**Keywords:** shield screw conveyor hydraulic control.

## 1 Introduction

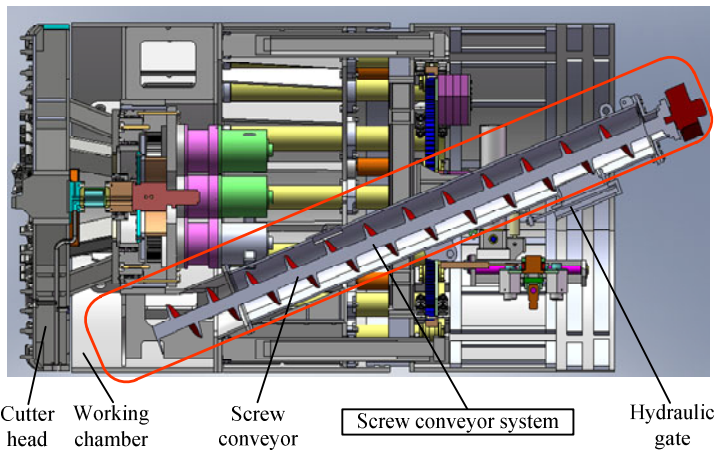
Shield tunnelling machine (shield for short) is a modern construction machinery dominating the tunneling field [1]. Characterized by quick and safe construction, high automation, and environment friendliness, it is beyond all question that the shield tunneling machine is the most promising and competitive equipment in underground space technology.

Screw conveyor system, driven by electro-hydraulic system, plays a fundamental role during the earth pressure balance (EPB) operations particularly for a correct application of the face pressure [2]. A screw conveyor extracts the soil from the head chamber, allowing control of the volume of soil excavated by the machine to balance the earth pressure in the head chamber [3]. Mainly composed of screw conveyor and hydraulic gate, which is positioned at the rear end of the conveyor [4], the screw conveyor system has three functions: muck disposal, water blocking by discharged muck sealed in the screw, keeping earth pressure balance by timely adjustments of rotation or opening on the basis of earth pressure monitoring. Fig. 1 shows the screw conveyor system in shield machine.

The present screw conveyor hydraulic system uses two pumps to supply pressure oil to screw motor and gate cylinder independently, which is rather complicated. For the sake of simplicity and economy of the system, just one variable displacement pump is used to supply pressure oil both to the motor and cylinder in the new system. However, the new screw conveyor hydraulic system is inherently compounded. The

compound internal relationship is adverse and undesirable, because the motor speed should be adjusted in accordance with the needs of EPB system rather than be affected by other disturbance. Hence, the major challenge involved in the proposed design of the control system is how to decouple the two actuators effectively. Without proper measures, the revolution speed of screw conveyor will be affected greatly regarding to the following simulation results.

This paper mainly deals with the disturbance compensation and control technology of the rotation and gate movement compound system. Firstly, a screw conveyor system with disturbance compensation and control technology is presented. It typically consists of a variable displacement pump, of which the displacement is adjustable according to the constant motor speed and the fluctuating of gate opening. Then the modeling, control design and dynamic simulations of the screw conveyor system are carried out. Finally, some beneficial conclusions are presented.



**Fig. 1.** The screw conveyor system in shield machine

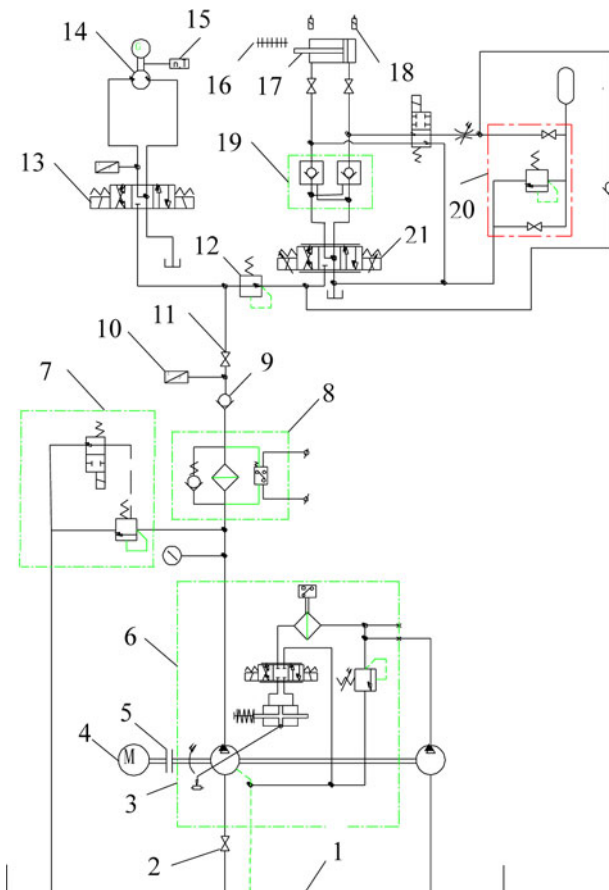
## 2 System Modeling

### 2.1 Screw Conveyor System

To prevent the occurrence of ground heave or settlement during shield excavating, the earth pressure balance must be achieved either by adjusting thrust speed to change intake volume or by adjusting screw conveyor speed to change discharge volume. According to [5], the thrust speed is generally determined based on geological conditions and should be set as high as reasonably possible in practice. That is to say, the earth pressure balance is more widely realized by controlling the discharge volume timely. Screw conveyors are widely used for transporting and/or elevating particulates at controlled and steady rates [6]. Control of the excavation process of an earth pressure balance (EPB) tunnelling machine depends on control of the soil flow through the

screw conveyor [7]. Hence, the electro-hydraulic system for screw conveyor must be able to adjust the motor speeds in real-time according to earth pressure.

Fig. 2 shows the hydraulic circuit diagram of the screw conveyor system in shield machine. As shown in this figure, the variable displacement pump 6 integrated with a proportional electromagnet supplies pressure oil to motor and cylinder, by which screw conveyor and gate are driven respectively. The displacement of the pump is regulated in accordance with the desired motor speed and cylinder velocity variation in the pump controlled motor system. In order to examine the rotation speed of the screw conveyor, an angular speed sensor 15 is applied. The measured earth pressure signal from the working chamber is transmitted to the proportional amplifier of the pump to adapt the output flow, thereby the desired rotation speed of screw conveyor or opening of gate is fulfilled [5]. The control block of the hydraulic motor driving system is illustrated in Fig. 3.



**Fig. 2.** Hydraulic circuit diagram of the screw conveyor system in shield machine

## 2.2 Mathematical Model

In pump displacement regulating process, based on the assumption that servo valve orifices are matched and symmetrical, the linearized servo valve flow equation is

$$q_l = k_q x_v - k_c p_c \quad , \quad (1)$$

Where  $q_l$  denotes the cylinder flow rate,  $k_q$  denotes the valve flow gain,  $x_v$  denotes the spool displacement,  $k_c$  denotes the valve flow-pressure coefficient, and  $p_c$  denotes the load pressure. The forward flow equation of the displacement change mechanism piston chamber is

$$q_1 = A_p \frac{dx_p}{dt} + C_{id}(p_1 - p_2) + C_{ed}p_1 + \frac{V_1}{\beta_p} \frac{dp_1}{t} \quad , \quad (2)$$

and the return flow equation of the piston chamber is

$$q_2 = A_p \frac{dx_p}{dt} + C_{id}(p_1 - p_2) - C_{ed}p_2 - \frac{V_2}{\beta_p} \frac{dp_2}{t} \quad , \quad (3)$$

Where  $A_p$  is the effective piston area,  $q_1$  is the forward flow rate,  $q_2$  is the return flow rate,  $x_p$  is the piston displacement,  $C_{id}$  is the internal leakage coefficient,  $C_{ed}$  is the external leakage coefficient,  $p_1$  is the forward chamber pressure,  $p_2$  is the return chamber pressure,  $V_1$  is the volume of forward chamber,  $V_2$  is the volume of return chamber, and  $\beta_p$  is the effective bulk modulus of system.

In order to simplify the dynamics analysis of the system, the load flow represented by the average of forward and return flow is described as

$$q_l = \frac{q_1 + q_2}{2} \quad . \quad (4)$$

Equations (2), (3), and (4) can be combined to yield

$$q_l = A_p \frac{dx_p}{dt} + C_{lp} p_c + \frac{V_t}{4\beta_e} \frac{dp_c}{dt} \quad , \quad (5)$$

where  $C_{lp}$  is the total leakage coefficient, and  $V_t$  is the total volume in both chambers. The dynamics equation of the cylinder is described as

$$A_p p_c = m_e \frac{d^2 x_p}{dt^2} + B_p \frac{dx_p}{dt} + kx_p \quad , \quad (6)$$

where  $m_e$  is total equivalent mass,  $B_p$  is the viscous damping coefficient, and  $k$  is the stiffness of the centering spring. If we confine ourselves to the vicinity of one particular operating point, the relationship between the piston displacement and pump control angle can be approximately described as

$$\gamma = k_\gamma x_p \quad , \quad (7)$$



where  $\gamma$  is the pump stroke angle and  $k_\gamma$  is the coefficient between piston displacement and stroke angle. The displacement of the pump is described as

$$D_p = k_p \gamma \quad , \quad (8)$$

where  $k_p$  is the displacement gradient of pump control. Therefore, the output flow rate of the pump is

$$q_{pout} = n_p D_p - C_{ip} (p_l - p_r) - C_{ep} p_l \quad , \quad (9)$$

where  $n_p$  is the rotation speed of the pump,  $C_{ip}$  is the internal leakage coefficient of the pump,  $C_{ep}$  is the external leakage coefficient of the pump,  $p_r$  is the pressure at the inlet of the pump approaching to zero, and  $p_l$  is the pressure at the outlet of the pump. The dynamic voltage equation for the coil is

$$u(t) = L \frac{di}{dt} + iR + k_v \frac{dx_v}{dt} \quad , \quad (10)$$

where  $u$  is the input voltage,  $L$  is the inductance,  $i$  is the current,  $R$  is the total resistances of the circuit, and  $k_v$  is the counter electromagnetic force constant. Because the current force characteristic of proportional solenoid is approximately linear, the output electromagnetic force can be written as

$$F_M = k_F i \quad , \quad (11)$$

where  $F_M$  denotes the electromagnetic force, and  $k_F$  denotes current force gain. The force equilibrium relationship of the valve spool is

$$F_M = m_v \frac{d^2 x_v}{dt^2} + D \frac{dx_v}{dt} + k_v x_v \quad , \quad (12)$$

where  $m_v$  is the mass of the spool,  $D$  is the coefficient of viscous friction, and  $k_v$  is the stiffness of the centering spring. The flow continuity equation for the main oil way is

$$q_{pout} = q_p + q_g \quad , \quad (13)$$

Where  $q_p$  is the flow rate at motor inlet, and  $q_g$  is the flow rate at cylinder inlet. The flow continuity equation for the forward chamber of the screw conveyor motor can be written as

$$q_p = C_{im} (p_l - p_m) + C_{em} p_l + D_m \frac{d\theta_m}{dt} + \frac{V_0}{\beta_m} \frac{dp_l}{dt} \quad , \quad (14)$$

where  $C_{im}$  is the internal leakage coefficient,  $p_m$  is the pressure at motor outlet approaching to zero,  $C_{em}$  is the external leakage coefficient,  $D_m$  is the volumetric displacement,  $\theta_m$  is the angular displacement,  $\beta_m$  is the effective bulk modulus, and  $V_0$  is the total volume of forward chamber. Assuming lumped constants to describe the load, the torque balance equation of screw motor is

$$D_m p_l = J_t \frac{d^2 \theta_m}{dt^2} + B_m \frac{d\theta_m}{dt} + T_l \quad , \quad (15)$$

where  $J_l$  is the total inertia moment of motor and load(referring to motor shaft),  $B_m$  is the total viscous damping coefficient, and  $T_l$  is the load torque.

Applying Laplace transformation to all above equations yields the transfer function of the control system. The block diagram of rotation and gate movement compound control system is shown in Fig. 3.

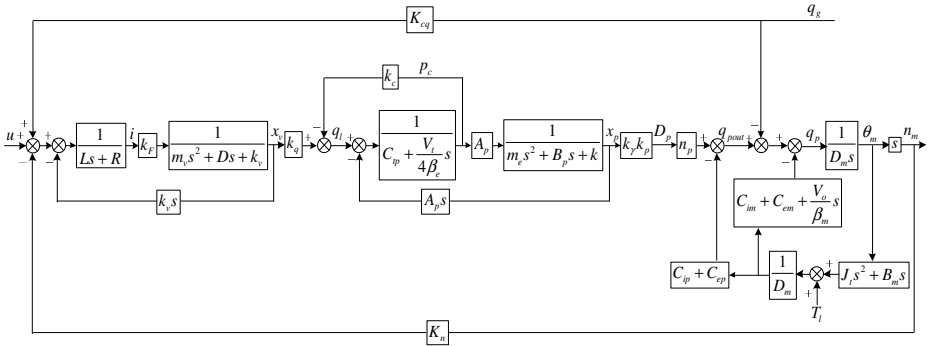


Fig. 3. Block diagram of rotation and gate movement compound control system

In the tunnel construction process, additives are often injected into the working chamber to improve the characteristics of the excavated soil so that it is plasticized and easy to convey [8]. Based on this assumption, the intake rate of the shield soil chamber is

$$q_i = \pi R^2 v \quad , \quad (16)$$

where  $R$  is the radius of cutter head and  $v$  is the thrust speed. The discharge rate by screw conveyor is

$$q_o = \pi \rho^2 T n_e \quad , \quad (17)$$

where  $\rho$  is the radius of the screw vane,  $T$  is the screw pitch, and  $n_e$  is the angular speed of screw conveyor determined by the EPB system. The flow continuity equation for plasticized earth will be

$$q_i = q_o + C_{eq} (p_e - p_o) + \frac{V_e}{\beta_e} \frac{dp_e}{dt} \quad , \quad (18)$$

where  $C_{eq}$  is the external leakage coefficient,  $p_e$  is the earth pressure in chamber,  $p_o$  is the leakage earth pressure approaching to zero,  $V_e$  is the volume of the chamber, and  $\beta_e$  is the bulk modulus of the stuffs in chamber.

Fig. 4 shows the screw conveyor speed regulating process in discharge-mode method of EPB control [9].Applying Laplace transformation to the above three equations yields the transfer function of speed setting system. The block diagram is illustrated in Fig. 5. The initial value of coil voltage is calculated with the set value of screw conveyor speed.

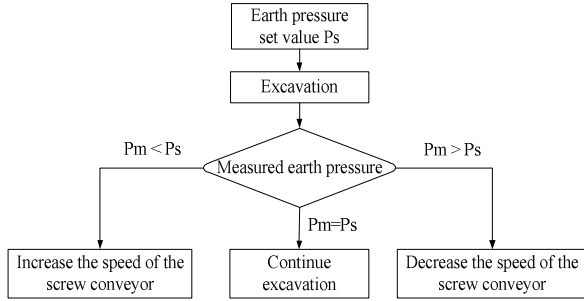


Fig. 4. Flow chart of discharge volume control

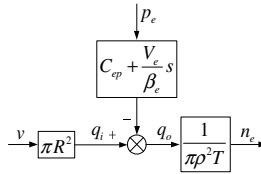


Fig. 5. Block diagram of the screw conveyor speed setting process

2.3 Control Design

In the practical excavating process, there are coupling relations between thrust velocity and screw conveyor angular speed. Once the thrust speed is set based on the geological conditions, the screw conveyor is also definite and should be maintained at a relatively constant value so as to keep the variation of shield chamber pressure within one acceptable range.

Fig. 6 shows the working principle of the compound control system. Both the motor and cylinder are driven by one pump, thus the desired motor speed will be affected by the fluctuating gate opening, if no measure is taken to compensate the flow rate variation.

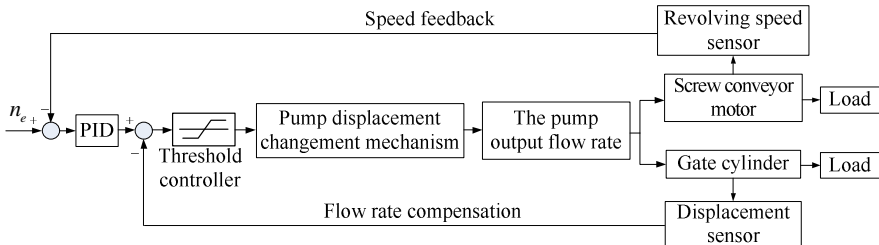


Fig. 6. Principle of rotation and gate movement compound control

As is shown in Fig. 4, if we regulate the input voltage of the pump displacement change valve, the angular speed of motor can be changed in the end. The voltage-flow characteristic of the pump is shown in Fig. 7 according to manufacturer's catalogue. The motor speed is measured with an angular speed sensor equipped inside the conveyor, and is converted into velocity parameter to be fed back so as to fulfill the outer loop. By comparing the reference and feedback signals, errors are produced and then sent to controllers after being transferred into electrical signals.

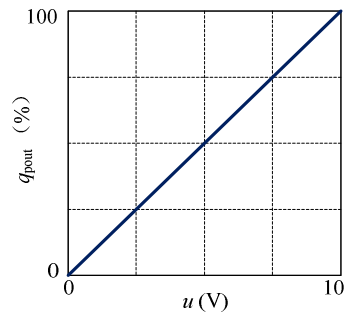
The gate opening is measured with displacement sensor equipped inside the cylinder, which is simultaneously converted into velocity parameter, and then the forward flow rate is calculated. Furthermore, the flow rate parameter is converted into voltage applied to the pump displacement control valve to increase or decrease the pump displacement to compensate the flow rate variation to fulfill the inner loop. In this way, the total flow rate varies in accordance with the cylinder flow rate, so the flow rate into the motor can be maintained at a constant value and a relatively stable motor speed can be obtained. The major innovative aspect of the approach utilized in the design of the control system is the double loop control algorithm. The inner loop works to have the control system make fast response to the gate movement, and the outer loop works to raise the system stable accuracy. Hence, both the dynamic and steady state performances are improved.

### 3 Simulation

The simulation is carried out in AMESim environment [10], as shown in Fig. 8. The model of the pressure reducing valve is conducted with the tool of HCD in the software while some other components having little relations to the control process are neglected. Here the load of the screw conveyor is represented with a revolving variable viscous model according to the soil mechanics [11]. The parameters involved in the simulation are set in accordance with the actual system and listed in Table 1, where  $v_g$  denotes the cylinder speed.

**Table 1.** Values of simulation parameters

Parameters	Values
$T_l$	10000Nm
$D_m$	1300ml/r
$n_{\text{emax}}$	24.6r/min
$n_{\text{emin}}$	10.5r/min
$v_{g\text{max}}$	1.5m/min
$v_{g\text{min}}$	1.1m/min



**Fig. 7.** Voltage-flow characteristic of the pump

In the simulation, we investigate the performance of the system under two sets of typical working conditions: a) Cylinder is at low speed while the motor works at low and high speed respectively, b) Cylinder is at high speed while the motor works at low and high speed respectively. Here, LCS, HCS, HMS and LMS are used to represent low cylinder speed, high cylinder speed, high motor speed and low motor speed respectively for short. The screw conveyor speed and cylinder speed parameters are illustrated in Table 1. Fig. 9 shows the simulation results of screw conveyor system under the first set of conditions, where  $n_{e1}$  is revolution speed with compensation and  $n_{e2}$  is revolution speed without compensation. Fig. 13 shows screw conveyor speed under the second set of conditions. It can be seen that the screw conveyor revolution speed is influenced severely by the hydraulic gate opening fluctuating if there is no speed compensation. Comparing the speed curves illustrated in Fig. 9 and Fig. 13, we can naturally derive that the proposed control strategy is able to perform the disturbance compensation function both under high and low speed conditions. As a result, the motor speed can be maintained at a relatively constant value. In the simulation, the gate opening begins to vary from the initial value at the 5th second. At the same time, the flow rate compensation mechanism begins to change the flow at the pump outlet. Fig. 10 and Fig. 14 illustrate the pump flow, motor flow and cylinder flow under the first and second sets of conditions respectively. The simulation results of cylinder flow and compensation flow ( $q_c$ ) under the four different conditions are shown in Fig. 11, Fig. 12, Fig. 15 and Fig. 16 respectively. By comparing the cylinder flow and compensation flow, the flow rate compensation effect is more clearly shown.

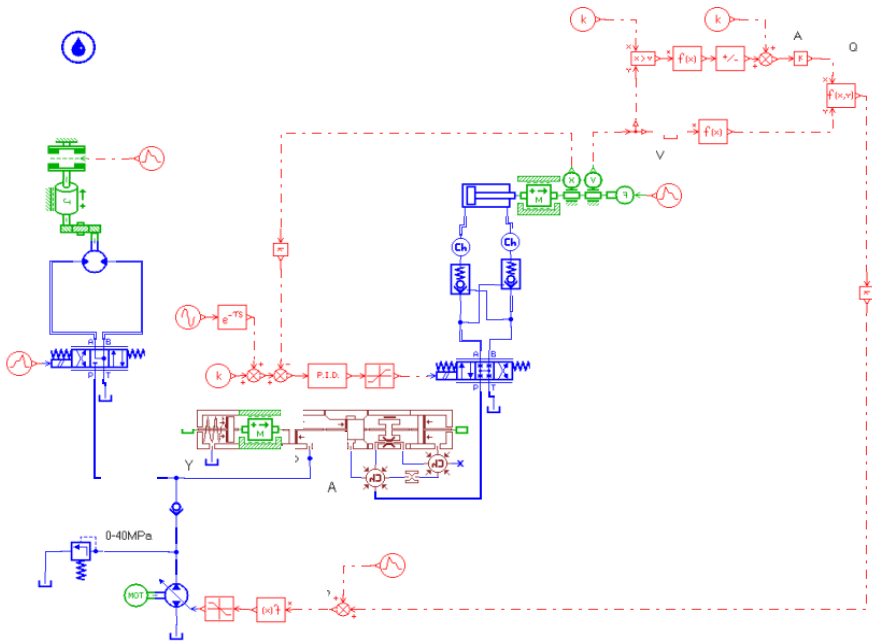


Fig. 8. Simulation model in AMESim

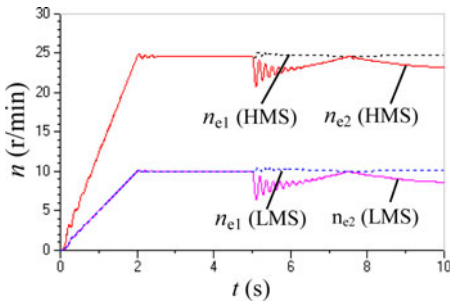


Fig. 9. Motor speed under LCS condition

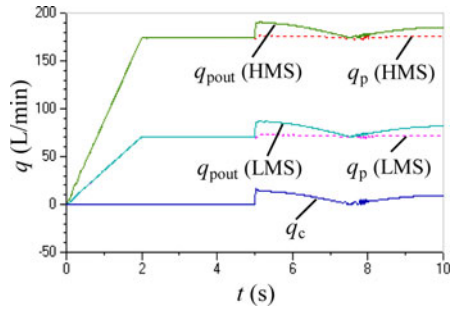


Fig. 10. Flow rate under LCS condition

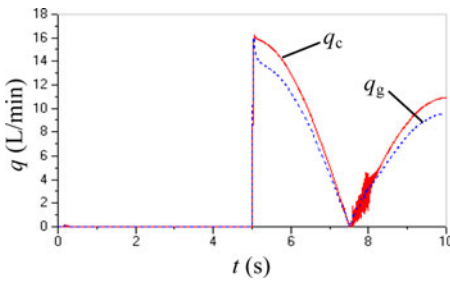


Fig. 11. Compensation flow HMS-LCS

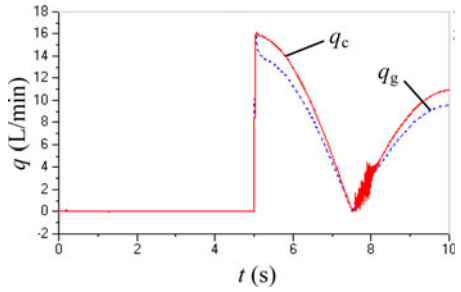


Fig. 12. Compensation flow LMS-LCS

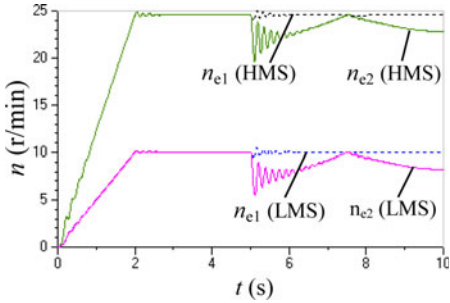


Fig. 13. Motor speed under HCS condition

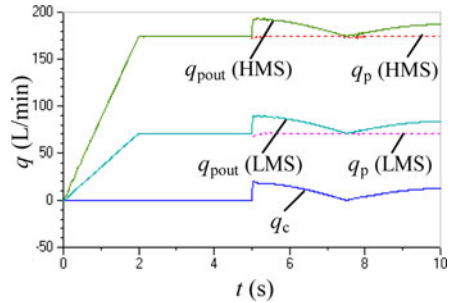


Fig. 14. Flow rate under HCS condition

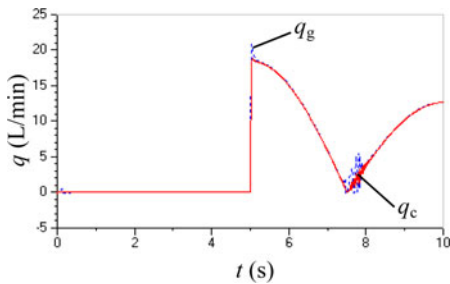


Fig. 15. Compensation flow HMS-HCS

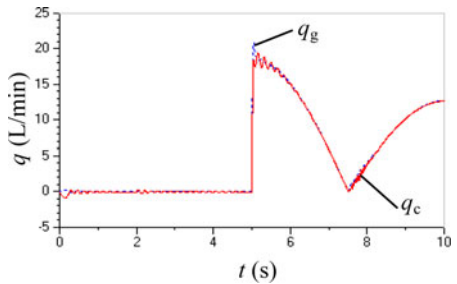


Fig. 16. Compensation flow LMS-HCS

## 4 Application

The electro-hydraulic control system with disturbance compensation has been successfully applied in the field application by now. This system has great superiority over the double-pump system in terms of structure simplicity and economy. The problem resulted from the undesired system internal coupling relationship has also been solved effectively with the compound control strategy. The realization of relatively stable screw conveyor speed provides a desirable precondition of the easier realization of earth pressure balance.

## 5 Conclusion

Based on the theoretical and simulation analysis, the following conclusions can be drawn.

- (1) The compensation flow is almost equal to the cylinder flow when the system is in equilibrium. Hence, flow rate compensation mechanism is functional.
- (2) The screw conveyor speed can be maintained at a relatively constant speed by adjusting the flow rate feedback coefficient under different work conditions.
- (3) Rotation and gate movement compound control of screw conveyor system with flow rate disturbance compensation is effective and the new screw conveyor system can satisfy the requirements of EPB system.

## Acknowledgments

This work is supported by National Basic Research Program (973 Program) of China (No.2007CB714004) and National High-tech Research and Development Program (863 Program) of China (No. 2007AA041806). Their support is gratefully appreciated.

The authors also wish to acknowledge the help of the engineers in Shenyang Heavy Machinery TBM Company.

## References

1. Huayong, Y., Guofang, G.: Shield TBM and its applications of hydraulic technology. *Hydraulics Pneumatics & Seals* 1, 27–29 (2004)
2. Daniele, P., Claudio, O., Raffaele, V.: Screw conveyor device for laboratory tests on conditioned soil for EPB tunnelling operations. *Journal of Geotechnical and Geoenvironmental Engineering* 133(12), 1622–1625 (2007)
3. Shouju, L., Zichang, S.G., Wei, S., Honhxia, D.: Controlling earth pressure of shield machine with optimization algorithm. In: *IEEE International Conference on Environmental Science and Information Application Technology*, Dubai, vol. 3(6), pp. 656–659 (2009)
4. Ning, C., Zhongzhi, S.: Design of the hydraulic system of shield's screw conveyor. *Journal of Zhejiang University of Science and Technology* 16(3), 164–166 (2004)

5. Huayong, Y., Hu, S., Guofang, G., Guoliang, H.: Earth pressure balance control for EPB shield. *Sci. China Ser. E-Tech. Sci.* 52(10), 2840–2848 (2009)
6. Owen, P.J., Cleary, P.W.: Prediction of screw conveyor performance using the discrete element method (DEM). *Powder Technology* 193(3), 274–288 (2009)
7. Merritt, A.S., Mair, R.J.: Mechanics of tunnelling machine screw conveyors: model tests. *Geotechnique* 56(9), 605–615 (2006)
8. Nomoto, T., Imaura, S., Hagiwara, T., et al.: Shield tunnel construction in centrifuge. *J. Geotech. Geoenviron. Eng.* 125(4), 289–300 (1999)
9. Hisakazu, M.: Earth pressure balanced shield method. *Tunnels & Tunnelling* 7(1), 47–49 (1980)
10. Hu, S., Guofang, G., Huayong, Y.: Pressure and Speed Control of Electro-hydraulic Drive Intelligent Mechatronics, Singapore, pp.314–317 (2008)
11. Guoliang, H., Guofang, G., Huayong, Y.: Thrust hydraulic system of shield tunnel boring machine with pressure and flow compound control. *Chinese Journal of Mechanical Engineering* 42(6), 124–127 (2006)



# Applications of the Overlapping Nodes Method in Simulating Crack-Expansion under the Action of TBM Cutter

Kui Zhang<sup>1</sup> and Yi-min Xia<sup>1,2</sup>

<sup>1</sup> College of Mechanical and Electrical Engineering, Central South University, Changsha Hunan, 410083, China

<sup>2</sup> Key Laboratory of Modern Complex Equipment Design and Extreme Manufacturing, Ministry of Education, Central South University, Changsha Hunan, 410083, China  
zhangkui414313@yahoo.com.cn, xiaymj@mail.csu.edu.cn

**Abstract.** In order to make good use of ANSYS/LS-DYNA, the overlapping nodes method is first applied to establish the TBM cutter's rock-breaking model. The modeling procedure is further discussed. By running the self-compiled program, the sharing node in rock's adjacent finite elements is automatically split into two nodes and then welded together by adding the failure constraint connection. Finally, by comparing and analyzing the example, the conclusion can be drawn that the 2D simulation model using the overlapping nodes method can better simulate the dynamic evolution process of rock crack, relative to the traditional modeling method. The shape of crack is fine and accords well with the experimental phenomena.

**Keywords:** the overlapping nodes method; TBM cutters; numerical simulation; crack expansion.

## 1 Introduction

The research on rock-breaking mechanism of TBM cutters has an important significance for prediction of TBM's performance and design of cutter heads. The rock-expansion mechanism under the action of TBM cutters is one of the most crucial problems in this research.

Numerical simulation can not only repeat the process of initiation and propagation of rock crack, but also can easily make quantitative analysis of crack system, so it is more popular than conventional experiment in this field. Compared with UDEC software based on distinct element method and some other based on meshless method, the explicit integral dynamics analysis program ANSYS/LS-DYNA has more material types and is more widely used in solving some dynamical problems [1-3].

Unfortunately, the adjacent elements in the rock model which is conventionally established in ANSYS/LS-DYNA share nodes [4]. The slip of neighbor elements doesn't obviously appear, that is to say, the inter-granular crack of rock model

can't be simulated. Simultaneously, in the rock-breaking process, the program may automatically delete the failing elements, so rock crack even has no time to survive before it can be observed by people.

In order to resolve this problem, the overlapping nodes method is added into the conventional modeling process in ANSYS/LS-DYNA. The automatic transformation from sharing nodes to overlapping nodes is realized by compiling the Matlab program and then the overlapping nodes are welded together. Finally, a computing example show how the model is established by the overlapping nodes method and the different effect between two methods.

## 2 Principle of the Overlapping Nodes Method

The finite element model conventionally created by ANSYS/LY-DYNA is called sharing nodes model in which adjacent elements share the same nodes (node No.3 in Fig. 1a). By using the overlapping nodes method, we split the sharing node into two overlapping nodes (node No.3 and node No.2003 in Fig. 1a) and then adds a failure constraint connection (red lines in Fig. 1a) to weld them together [5]. When TBM cutter is cutting the rock, if the connection between the overlapping nodes fails, a crack is created and then will spread rapidly along the elements. The application of the overlapping nodes method in 2D model is essentially the same as in 3D model.

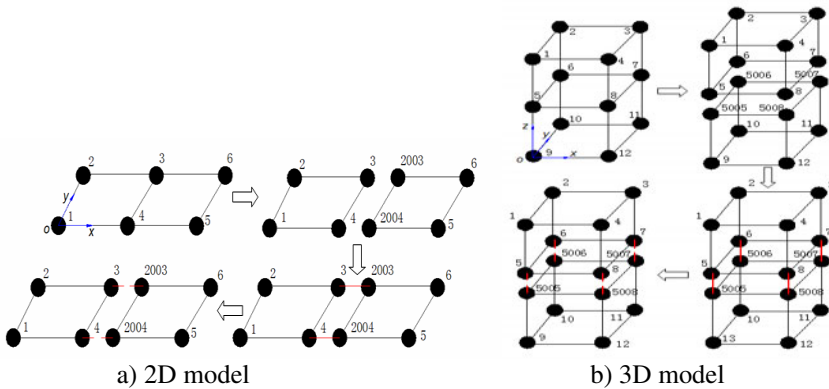


Fig. 1. Schematic diagram of the overlapping nodes method

What we should pay attention to is that if the rock material doesn't add additional failure pattern, the elements with the failing connection will just break but not be deleted. In this case, the rock mesh will become too contorted to continue the procedure by the action of cutter. In order to solve this problem, \*MAT\_ADD\_EROSION card can be used to add another failure criteria such as the biggest operation time of computing time and the largest main strain and so on.

### 3 The Application in Numerical Analysis

#### 3.1 The Modeling Procedure

As shown in Fig. 2, a 2D geometric model which consists of a TBM cutter and a piece of rock is initially created by PRO/E and then introduced into ANSYS. Mesh it and generate the original finite element model in an ordinary way. A list file (\*.lis) containing the coordinate information of nodes and the information of elements is exported by ANSYS. Delete the rock model and run the self-compiling program which uses the list file as input. An APDL command file (ANSYS.txt) which could produce the overlapping nodes automatically and rebuild the finite element of rock model in ANSYS is automatically generated. A key file (KEY.txt) which could be copied to the corresponding position in k file (\*.k) and complete the final modeling process by welding the overlapping nodes is also obtained.

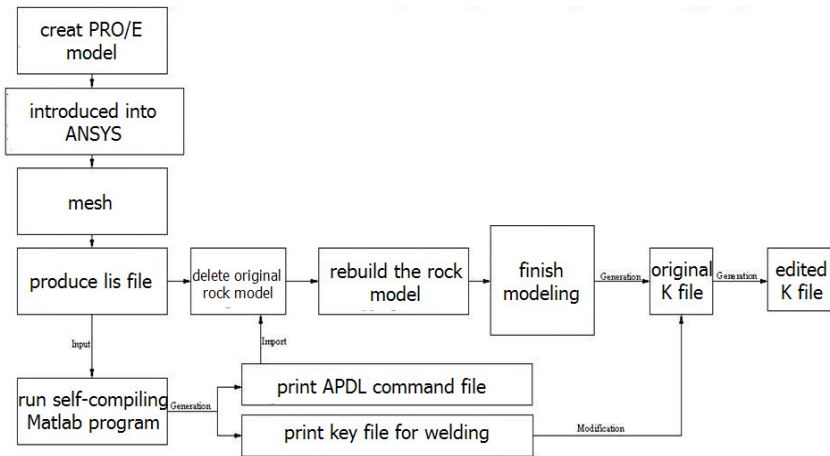


Fig. 2. Modeling diagrams

#### 3.2 The Self-compiling Program

Referred to Fig. 3, we have compiled a Matlab program to achieve the goal of automatic transformation.

#### 3.3 Data File Created by the Self-compiling Program

Taking the 2D model (as shown in Fig. 1a) for example, let the element length be 0.005mm. ANSYS.txt file is shown in Fig. 4a. KEY.txt file is shown in Fig. 4b.



## 4 Numerical Example Analysis

### 4.1 Introduction of Numerical Examples

For simplicity, this paper only introduces the application in 2d modeling of rock breaking by a standard TBM cutter. Choosing the concrete as a rock material, the size of the specimen is  $0.03m \times 0.05m$ , and the model scale is 1:1.

### 4.2 The Establishment of Finite Element Model

As shown in Table 1, the material type of rock is set as JOHNSO-HOLMGUIST [6]. The main parameters of the specimen are listed in following table. Other material parameters of model are based on the experimental data [7]. MAT\_ADD\_EROSION is used to add modes of additional tensile stress of concrete failure in order to prevent the extreme distortion of the grid in simulation process. Use the material type RIGID to simulate TBM cutter. Other parameters and element attributes are set as Table 2.

**Table 1.** Material parameters of rock

Density [kg/m <sup>3</sup> ]	2440
Uniaxial Compressive Strength of quasi-static [MPa]	48
Shear modulus [GPa]	14.86
Maximum static tensile stress [MPa]	4

**Table 2.** ANSYS material model and element attributes

Project	Element Type	Element Option	Element Size (mm)	Element Number	Material Model
disc-cutter	PLANE 162	Plane stress	1	560	*MAT_RIGID
Concrete	PLANE 162	Plane stress	0.1	150000	*MAT_ADD_EROSION *MAT_JOHNSON_HOLMGUIST_ CONCRETE

\*CONTACT\_2D\_AUTOMATIC\_SINGLE\_SURFACE is used as the contact algorithm between TBM cutter and rock. The cutting speed of the cutter is set to 0.1m/s, and the cutting depth is 0.007m. The final models are shown in Fig. 6.

### 4.3 Numerical Analysis Results

The numerical result by using the conventional modeling methods is shown as follows in Fig. 2.

The simulation result obtained by the overlapping nodes method is shown in Fig. 6. In the process of the initial invasion, the rock at the bottom edge of TBM cutter has deformed under the enormous compressive stress. The elements are removed by program when the distortion has reached the preset value. In Fig. 6a the cone failure appears at the bottom edge while the end region has generated a number of small Hertz cracks. Micro-cracks at the edge of conical failure zone tend to spread outward. Failure zone is under the high confining pressure conditions. Therefore the destruction does not occur, as shown in Fig. 6b. In Fig. 6c, primary radial and middle cracks continue to propagate while some cracks formed a new branch of bifurcation at the same time. By the overlapping nodes method, we can get precise simulation of a series of acts of expansion and bifurcation the generation of rock propagation.

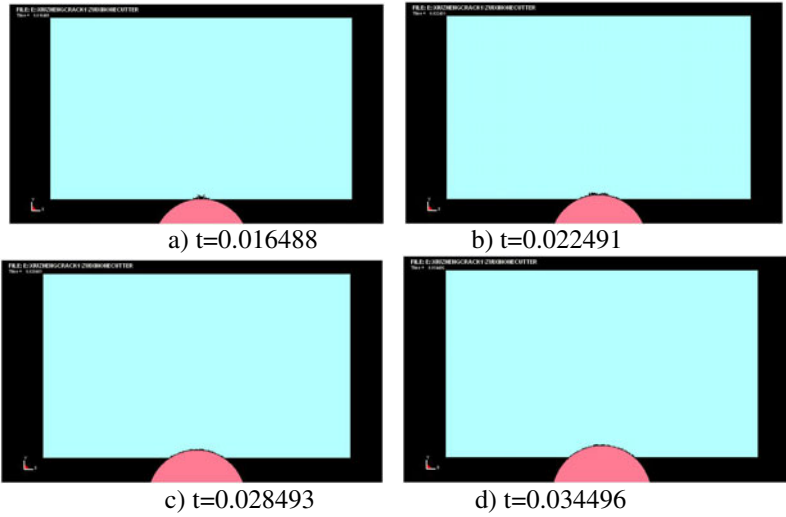


Fig. 5. Simulation results obtained by using the conventional modeling method

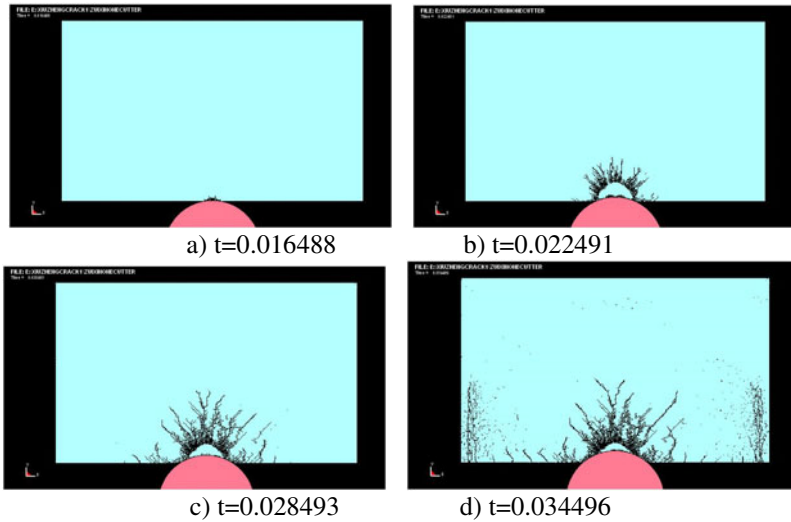


Fig. 6. Simulation results obtained by the overlapping nodes method

To sum up, we can draw the conclusion that the effect of the conventional modeling method is not as obvious as the effect of the overlapping nodes method under the same precision of element grids. The conventional modeling method requests for more precise element grid. Lindqvist, who did series of experiments in 1994, had observed and recorded the real evolution of the crack [8]. It is easy to say that the simulation result is consistent with the experimental phenomena by comparison.

## 5 Conclusion

The 2D model established by the overlapping nodes method can better simulate the dynamic process of crack production and expansion at the same precision level. Crack patterns are consistent with the experimental phenomena. However, there are still some defects in this method. For example, the self-compiling program will consume a great deal of computer resources. This method needs to be further improved.

## Acknowledgements

This research is supported by the National Key Basic Research and Development Program of China (973 Program, No. 2007CB714002) and Central South University Postdoctoral Sustentation Fund (743410151), which are gratefully acknowledged.

## References

1. Gong, Q.M., Jiao, Y.Y., Zhao, J.: Numerical modeling of the effects of joint spacing on rock fragmentation by TBM cutters. *Tunneling and Underground Space Technology* 21, 46 (2006)
2. Kou, S.Q., Lindqvist, P.A., Tang, C.A., Xu, X.H.: Numerical simulation of the cutting of inhomogeneous rocks. *Int. J. Rock. Mech. Min. Sci. Geomech. Abstr.* 36, 711 (1999)
3. Tan, X.C., Kou, S.Q., Lindqvist, P.A.: Application of the DDM and fracture mechanics model on the simulation of rock breakage by mechanical tools. *Engineering Geology* 49, 277 (1998)
4. Tan, Q., Tu, C.-F., Ji, Z.-Y., Xia, Y.-M.: Numerical Simulation of Rock Cutting by Disc CutteR 2, 62 (2009)
5. Bai, J.-Z.: Theoretical basis and example analysis on LS-DYNA3D. Science Press, Beijing (2005)
6. Hanchak, S.J., Forrestal, M.J., Young, E.R., Ehrigott, J.Q.: Perforation of Concrete Slabs with 48MPa (7ksi) and 140MPa (20ksi) Unconfined Compressive Strengths. *International Journal of Impact Engineering* 12, 1 (1992)
7. Holmquist, T.H., Johnson, G.R., Cook, W.H.: A computational constitutive model, for concrete subjected to large strains, high strain rates, and high pressures. In: Fourteenth International Symposium on Ballistics (1995)
8. Lindqvist, P.A., Lai, H.H., Alm, O.: Indentation fracture development in rock continuously observed with a scanning electron microscope. *Int. J. Rock mech. Min. Sci. Geomech. Abstr.* 21, 165 (1984)

# The Simulation Analysis of the TBM Disc Cutter Phase Angle

Kai Wang<sup>1</sup>, Qing Tan<sup>1</sup>, Yimin Xia<sup>1,2</sup>, and Zijun Xu<sup>1</sup>

<sup>1</sup> College of Mechanical and Electrical Engineering, Central South University, Changsha Hunan, 410083, China

<sup>2</sup> Key Laboratory of Modern Complex Equipment Design and Extreme Manufacturing, Ministry of Education, Central South University, Changsha Hunan, 410083, China  
wangkai00112@163.com, tanqing@mail.csu.edu.cn

**Abstract.** A series of numerical experiments was performed to simulate rock fragmentation by TBM (Tunnel Boring Machine) disc cutter. We used the large non-linear dynamic finite element software ANSYS/LS-DYNA to simulate three-dimensional dynamic process of two disc cutters cutting soft rock. The information such as when disc cutters combined, the fore disc cutter's has effect on the rear disc cutter under different cutting phase angle, and the soft rocks stress field of the disc cutters at its corresponded load steps and so on are obtained, what's more, the results had been carefully analyzed. The simulation results shows that, with the increase of phase angle of the two adjacent disc cutters, there is little change in the average lateral force and the average rolling force of the rear cutter. When the phase angle equals to 120°, the average vertical force obtains its minimum, and the maximal external lateral force of the rear cutter is larger than its maximal internal lateral force, then the lateral unbalance force points to inside. This conclusion provides basic data for the structural design of cutter head and engineering construction maintenance. The conclusion can be used as a guide for the further research of TBM cutters' rock fragmentation mechanism and can provide a reliable foundation for the improvement of the design of TBM cutters.

**Keywords:** TBM; soft rock; disc cutter; phase angle.

## 1 Introduction

Reasonable exploitation and utilization of underground space have been gradually recognized as the best way to promote the sustainable development of society and protect environment by both at home and abroad nowadays, which can solve the increasingly shortage of the construction land and the land resources. The TBM construction has the advantages of good control groundwater seepage and surface subsidence, no influence of underground water and ground transportation, construction and project cost controlled, relatively small construction risk, so it is widely used in the construction of subway tunnels, drainage tunnels and large cross-river tunnels both at home and abroad[1-2]. In the process of cutting rock, reasonable arrangement of cutters for best cutting performance is essential to the continuous effective excavation of



shield turning machine. And it is the key technology for excavation cost reduction. The selection of an appropriate cutters phase angle is not only the major factor of the cutter head’s dynamic balancing, but it is also important parameter for the cutting efficiency of cutters[3-5].

Literature [6-12] makes the simulation of the disc cutter cutting rocks and numerical simulation analyses of the disc cutter breaking rocks with different methods separately. However, little research has been conducted on the phase angle of TBM disc cutters presently. Therefore, in this paper, we choose appropriate soft rock constitutive model, employ the explicit dynamics analysis software ANSYS/LS-DYNA to make numerical simulation in the cutting soft rock process, researches the fore cutter’s influence on the rear cutter under different cutting phase with combination of cutters and the soft rock’s stress field of both the disc cutters at its corresponded load steps. The result can provide basic data for the structural design of cutter head and engineering construction maintenance and can be used as a guide for the further research of TBM cutters’ rock fragmentation mechanism and provide reliable foundation for the improvement of the design of TBM cutters.

## 2 Material Constitutive Model

In the numerical simulation of the cutting soft rock process, it is particularly critical to determine the material constitutive relation. In this study, we adopt the \*MAT\_JOHNSON\_HOLMGUIST\_CONCRETE material model in the LS-DYNA970. It is an elastic-plastic constitutive model introducing damage factor based on Ottosen four parameters failure criteria [13], and its constitutive model is

$$\sigma^* = [A(1 - D) + BP'^N] \left[ 1 + c \ln(\dot{\epsilon}^*) \right] \tag{1}$$

In the formula:  $D$  stands for damage factor.  $P' = P / f_c'$ ,  $P$  stands for actual pressure,  $f_c'$  represents quasi-static uniaxial compressive strength.  $\dot{\epsilon}^* = \dot{\epsilon} / \dot{\epsilon}_0$ ,  $\dot{\epsilon}$  stands for the actual strain rate, and  $\dot{\epsilon}_0$  is the reference strain rate.  $A$  stands for the rock dimensionless viscous constants,  $B$  stands for dimensionless pressure coefficient of strengthening,  $c$  stands for strain rate coefficient, and  $N$  stands for dimensionless pressure hardening exponent.

$$D = \sum \frac{\Delta \epsilon_p + \Delta \mu_p}{D_1 (P'' + T'')^{D_2}} \tag{2}$$

In the formula:  $\Delta \epsilon_p$  represents the equivalent plastic strain,  $\Delta \mu_p$  is the equivalent plastic volume strain.  $T'' = T / f_c'$ ,  $T$  stands for the maximum static material tensile stress,  $D1$  and  $D2$  stands for material damage constants.

$P$  stands for pressure; it is the volumetric strain function (including permanent collapses state). According to the rock fracture damage constitutive model, the

relationship between stress and volume strain can be divided into three stages, as is shown in figure 1.

The first stage is linear elastic stage

$$K_{elastic} = \frac{P_{crush}}{\mu_{crush}} \tag{3}$$

In the formula:  $P_{crush}$  and  $\mu_{crush}$  stand for yield stress value and volume strain value in the uniaxial compression test.

The second stage is transitional stage. In this stage the air is gradually pressed out from rock gap, so it produces plastic volume strain, using interpolation method to calculate stress:

$$\frac{p - P_{lock}}{P_{lock} - P_{crush}} = \frac{\mu - \mu_{lock}}{\mu_{lock} - \mu_{crush}} \tag{4}$$

In the formula:  $p_{lock}$  and  $\mu_{lock}$  stand for the pressure and volume strain of the material compaction point.

The third stage is close-grained stage, considering  $p$  is the function of volumetric strain  $\mu$ , namely:

$$p = k_1 \bar{\mu} + k_2 \bar{\mu}^2 + k_3 \bar{\mu}^3 \tag{5}$$

In the formula:  $\bar{\mu} = (\mu - \mu_{lock}) / (1 + \mu_{lock})$  for modified volumetric strain,  $\mu = \rho / \rho_0 - 1$  stands for standard volumetric strain,  $\rho$  stands for the current density of rock,  $\rho_0$  stands for the initial density of rock,  $k1$ ,  $k2$ ,  $k3$  stand for material constants.  $\mu_{lock} = \rho_{grain} / \rho_0 - 1$ , at this time, the air in the rock is all pressed out of the gap, and  $\rho_{grain}$  stands for dry density of rock.

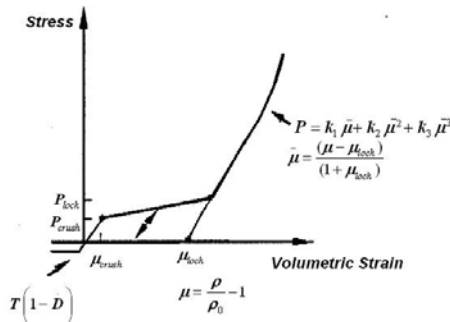


Fig. 1. Relationship of pressure and volumetric strain

### 3 Numerical Model

#### 3.1 The Building of Finite Element Model

For the observed certain type of TBM's cutter head, the phase angle of the layout in the disc cutter ranged from  $25.3^\circ$  to  $178.6^\circ$ . We developed the following simulation experiments with the following phase angle:  $60^\circ$ ,  $90^\circ$ ,  $120^\circ$ ,  $150^\circ$ ,  $180^\circ$ . Figure2 is the global graph when the phase angle difference is  $60^\circ$ . Establish the three-dimensional finite element model of the disc cutter at 1:1 size ratio. Sector cement concrete specimen thickness is 0.06m. And inter radius is 1.929m, the outer radius 2.071m respectively. The fan-shaped angle is  $10^\circ$ . The cutter body intimately contact with the cutter ring so that the cutter can drive the cutter ring to cut the rock. Considering the connection of accuracy and computational time, we adopt the solid164 hexahedron- eight node elements to mesh generation. The model showed in Figure3.

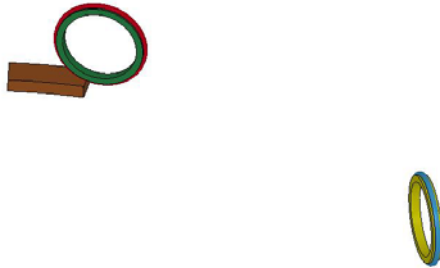


Fig. 2. The global graph when the phase angle is  $60^\circ$

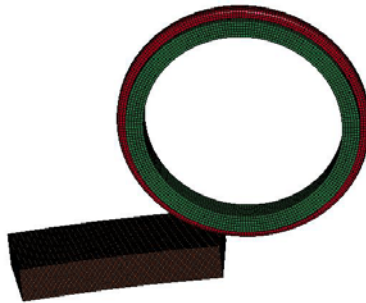


Fig. 3. The finite element model of the rock's fracture by disc cutter

In the simulating process, we choose rigid body model to describe the cutter ring, and cutter body described by elastic-plastic body. The specific parameters: density  $7850\text{Kg/m}^3$ , elastic modulus  $210\text{GPa}$ , Poisson's ratio 0.3. The main properties of soft rock model list: density  $2240\text{kg/m}^3$ , hear modulus  $12.61\text{GPa}$ , the static yield strength of  $37.1\text{MPa}$ , maximum principal stress  $3\text{MPa}$  in stretching mode.

### 3.2 The Determination of the Construction Parameters

Choose disc cutter's cutting depth to be 8mm, cutter head's rotational speed to be 6r/min. In the study, two inner cutters are fixed 2m from the center of the cutter head. Thus the revolution angular rate of the disc cutter is 0.628rad/s.

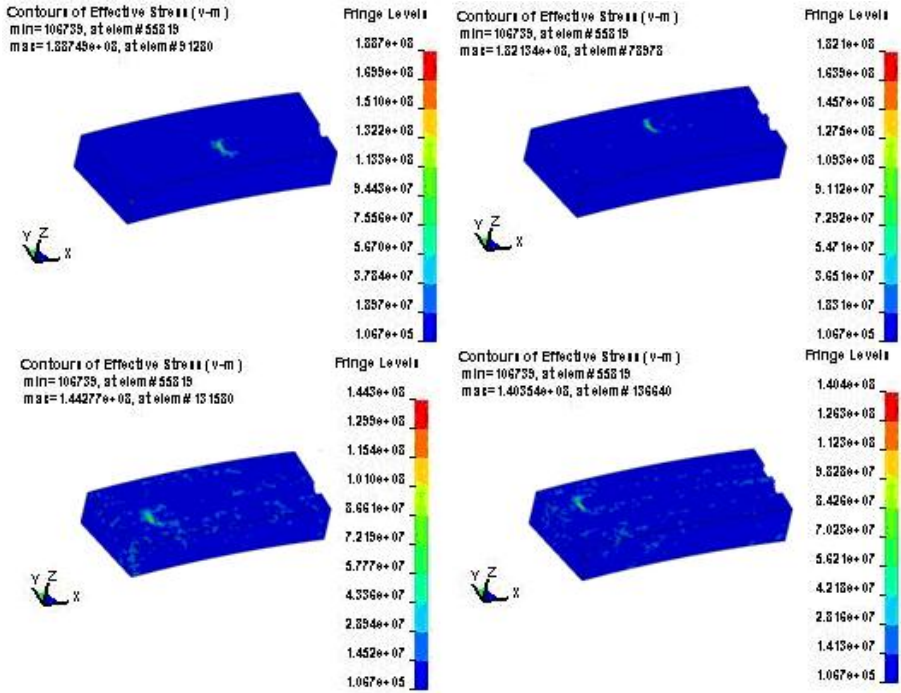
The contact of the cutter with the soft rock can be seen as area contact of the rigid body with the soft body. In the rock-cutting process, severe nonlinear failure occurred in the soft rock. Continuously, the soft rock units fail and new soft rock units start to contact with the disc cutter. The purpose of the adoption of the erosion contact model is to assure that when those failed elements was removed, the rest elements can still contact with the disc cutter. We use penalty function to describe the eroding contact model in the simulation.

We choose the coefficient of friction to be 0.2 considering the friction of the cutter with the rock. And the maximum of the friction should be restricted in the simulation to avoid the unreasonable result. Set the viscous friction coefficient  $VC = \sigma_0 / \sqrt{3}$ ,  $\sigma_0$  is the yielding stress of the rock material.

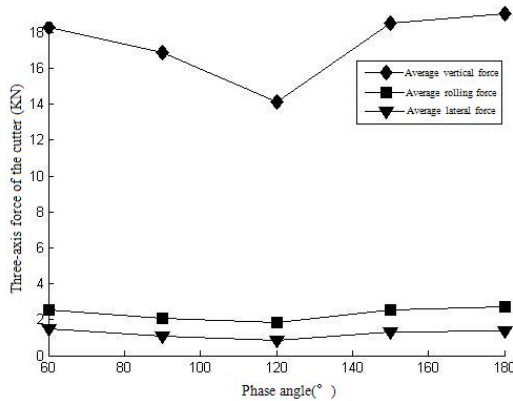
## 4 Simulation Results and Analysis

Figure 4 shows stress distribution of the soft rock in the process of the fore and rear cutter's cutting soft rock. When equivalent stress of the rock is over the maximum tensile strength set by the rock material' failure criterion, the rock will be pulled apart. As the rock was pulled, the rock bits was fractured from the contact department of cutting edge; and then the rock will be fractured, the elements lose effectiveness thus be removed. As a result, the rock material is divided into two parts, one is what has been cut, the other is the remains not. From the Figure4, we can conclude that when the second disc cutter cut the soft rock, the maximum stress is obviously less than that when the first disc cutter cut in the corresponding load step. Therefore, it is easier to fracture the rock after the rolling process.

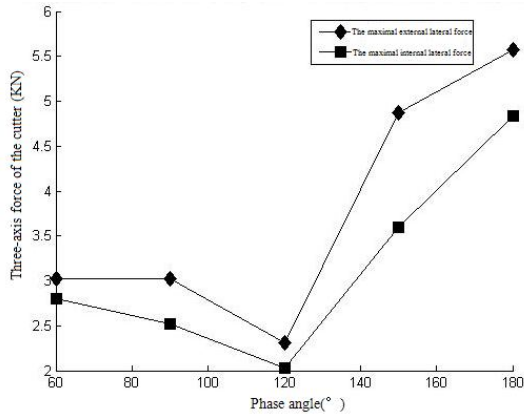
According to the simulating result, we can figure the average three-direction forces and the maximum medial and lateral forces of the rear disc cutter. They are showed in the Figure5 and Figure6. In the rolling process, there is an unbalanced lateral force whose force volatility can cause the cutter head lateral loose, result in the circlip displacement or fracture failure. Further, it will cause the lateral movement of the cutter ring and the lateral imbalance force. The rear disc cutter is subjected to a bigger maximum external lateral force than the maximum internal lateral force and the composite force points to the medial. With the increase in phase angle of two adjacent disc cutters, there is not much difference in the average lateral force and the average rolling force of the rear cutter. What we can conclude is that the average vertical force achieved the minimum value when the phase angle is 120°.



**Fig. 4.** The equivalent stress of the rock in corresponding forces. The first row on the left: The eighth load step of the fore cutter; The first row on the right: The eighth load step of the rear cutter; The second row on the left: The thirteenth load step of the fore cutter; The second row on the right: The thirteenth load step of the rear cutter.



**Fig. 5.** The average three-direction forces of the disc cutter when double cutters break the rock



**Fig. 6.** The maximum medial and lateral forces of the disc cutter when double cutters break the rock

## 5 Conclusions

(1) A finite element dynamic model based on ANSYS-LS/DYNA software is established to simulate the rock fracture process when the rock is forced by two disc cutters in different phase angle. The result is able to reflect the real process.

(2) The equivalent stress of the rock in different load step is obtained. In the corresponding load step, when the second disc cutter cutting the soft rock, the maximum stress is obviously less than the first disc cutter.

(3) The average three-direction forces and the maximum medial and lateral forces have been obtained when disc cutters at different phase angle cut the rock one after another. And we summarized the regulations to provide the basis to arrange the disc cutter in the cutter head.

## Acknowledgement

This research is supported by the National Key Basic Research and Development Program of China (973 Program, No. 2007CB714002) and Central South University Postdoctoral Sustentation Fund (743410151), which are gratefully acknowledged. The authors would like to thank Mr. YI Nian-en, and Ms. LV Dan, for the useful suggestions during the preparation of this paper.

## References

1. Guo-hua, C., Guo-qiang, W., En-guang, H.: The research status and prospect of shield machine. *Mining Machinery* 34(6), 24–27 (2006) (in Chinese),
2. Eun-Soo, H., Lee, I.-M., Shin, H.S.: Quantitative risk evaluation based on event tree analysis technique: Application to the design of shield TBM. *Tunnelling and Underground Space Technology* 24(3), 269–277 (2009)

3. Cho, J.-W., Jeon, S., Yu, S.-H.: Optimum spacing of TBM disc cutters: A numerical simulation using the three-dimensional dynamic fracturing method. *Tunnelling and Underground Space Technology* 25(3), 230–244 (2010)
4. Chang, S.-H., Choi, S.-W., Bae, G.-J., Jeon, S.: Performance prediction of TBM disc cutting on granitic rock by the linear cutting test. *Tunnelling and Underground Space Technology* 21(3), 271 (2006)
5. Jun-zhou, H., Yan-jun, S., Hong-fei, T.: The design method of TBM cutters layout. *China Mechanical Engineering* 19(15), 1832–1836 (2008)
6. Gertsch, R., Gertsch, L., Rostami, J.: Disc Cutting Tests in Colorado Red Granite: Implications for TBM Performance Prediction. *International Journal of Rock Mechanics and Mining Sciences* 44(2), 238–246 (2007)
7. Jian-min, G., Peng, Z., Bing, Z.: Development and Test of High Speed Soil-Cutting Simulation System Based on Smooth Particle Hydrodynamics. *Transactions of the CSAE* 23(8), 20–26 (2007)
8. Abo-Elnor, M., Hamilton, R., Boyle, J.T.: Simulation of soil-blade interaction for sandy soil using advanced 3D finite element analysis. *Soil & Tillage Research* (75), 61–73 (2004)
9. Jun-hong, D., Xian-long, J., Yi-zhi, G.: Study on 3D Numerical Simulation for Soil Cutting with Large Deformation. *Transactions of the Chinese Society for Agricultural Machinery* 38(4), 118–121 (2007)
10. Gong, Q.-M., Zhao, J., Jiao, Y.-Y.: Numerical modeling of the effects of joint orientation on rock fragmentation by TBM cutters. *Tunnelling and Underground Space Technology* (20), 183–191 (2005)
11. Gong, Q.M., Zhao, J., Hefny, A.M.: Numerical simulation of rock fragmentation process induced by two TBM cutters and cutter spacing optimization. *Tunnelling and Underground Space Technology* 21(3), 263 (2006)
12. Jia-ning, S., Qin, L., Li, T.: Emulation of the Rock Fragmentation Principle of the Percussion N-Rotary Drill Bit. *Journal of South west Petroleum University* 11(29), 113–115 (2007) (in Chinese)
13. Holmquist, T.H., Johnson, G.R., Cook, W.H.: A computational constitutive model, for concrete subjected to large strains, high strain rates, and high pressures. In: *Fourteenth International Symposium on Ballistics*, pp. 591–600 (1995)

# The Design of Shield Tunnelling Machine Cutter Header's Cutting Performance Test Bed

Zongming Zhu<sup>1,2,\*</sup>, Yimin Xia<sup>1,3</sup>, Dezhi Luo<sup>1</sup>, and Tao Teng<sup>1</sup>

<sup>1</sup> Key Laboratory of Modern Complex Equipment Design and Extreme Manufacturing, Ministry of Education  
410083, Changsha, Hunan, China  
Tel.: +86 571 84261492

<sup>2</sup> The Department of Mechanical and Electronic Engineering, Changsha University  
410003, Changsha, Hunan, China

<sup>3</sup> School of Materials Science and Engineering, Central South University  
410083, Changsha, Hunan, China

zhuzm1977@163.com, xiaymj@mail.csu.edu.cn, luodz@126.com,  
38068914@qq.com

**Abstract.** In order to solve the problems of data access and low accuracy on the shield tunnelling machine construction site, a test bed that can simulate the shield tunnelling machine cutter head's cutting performance is designed. The test bed's box uses a fixed cutter head and a turning material bin. The whole configuration is designed to a welding framework. The cutter head is designed to a spoke form. The fixing angle and the distance between different cutters can be adjusted at random. The propelling and loading system of the test bed are controlled in an electro-hydraulic proportional control manner. It can be used to test the single factor and multiple factors that influence the cutter head's cutting performance.

**Keywords:** shield tunnelling machine; test bed; cutter head; hydraulic system.

## 1 Introduction

The shield tunnelling machine is a kind of construction machinery that widely used which find wide use in tunnel construction of city subways, railways, highways and water conservancy. It can realize the tunneling driving process factory production of geotechnical cutting, slag discharge, tunnel supporting, the control of ground settlement and so on [1]. The cutter head is an important part of the shield tunnelling machine, whose function is cutting the soil. Because the process of cutting the soil is complex and the circumstance is tough and the testing instruments' requests for circumstance and worksite are restricted, it is always difficult to solve the field measurement of shield tunneling machine cutter heads. At present, foreign companies such as Mitsubishi, Herrenknecht, several domestic universities and companies have already working in this field, and developed some different functional test devices. By laboratory simulation, scholars have done many research works in the hydraulic

---

\* The Corresponding author.



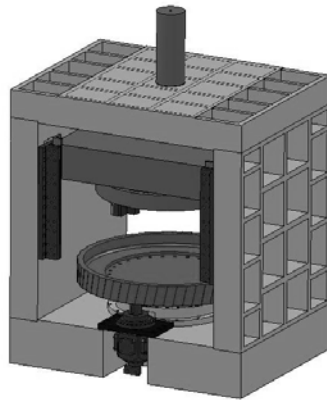
control, attitude adjustment, tunneling process excavation process, the design of the cutter head [2-4]. But the current test apparatus have certain limitations, and the rock breaking mechanism and the cutter head design are little researched, it is an urgent need to develop a more advance shield tunnelling machine test device, which provides the experimental basis for the research of rock breaking mechanism, cutter head's cutting performance and load character. Therefore we can deeply explore the single factor and multiple factors mechanism that influence cutter head's cutting performance. It is very important practical significance and strategic significance, that it has great realistic significance and strategic significance for solving the contradiction between domestic market demands for large-scale excavation equipments and the relatively hysteretic technical level, and enhancing our country's independent innovation capability and core competitiveness in equipment manufacturing industry.

So we designed a test bed with novel structure, small occupation of land, better overall stability, flexible assembly and adjustable cutter space and phase angle. Keeping all the above aspects in mind, a test rig is designed as a part of ongoing research to study the cutter head's cutting performance. The test bed has novel structure, small occupation of land, better overall stability, flexible assembly and adjustable cutter space and phase angle.

## 2 The Structure Design of Test Bed

The test bed overall dimension is 2220×1550×3100mm, and total weight is about 17 tons. The test bed's three-dimensional structural schematic diagram is showed in the following figure 1. The external structure of the test bed is mainly welded frame. The body is made up of a upper beam that fixed in the stent's upside and a lower beam that fixed in the beam's downside. The hydraulic cylinder is installed inside the upper beam, and the lower beam is fixed by anchor bolts. The cutter head is fixed together with the slider, and the piston rod of the hydraulic cylinder pushes the cutter head and the slider to straightly move up and down.

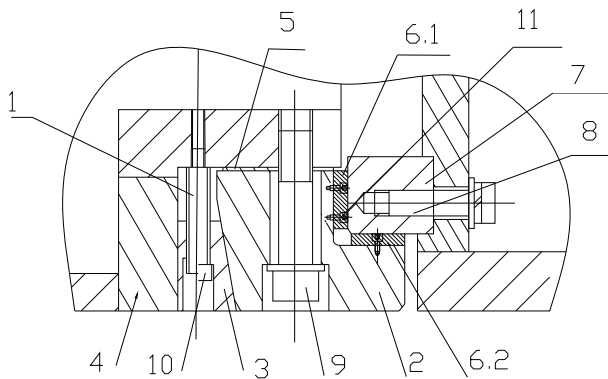
A guide device is distributedly installed in the inner side of the slider's four corner in the test bed body. The body adopts the inner wedged type four corner and eight



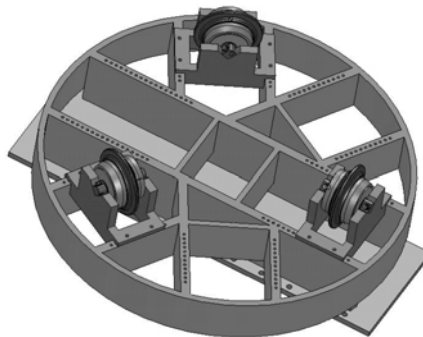
**Fig. 1.** Three-dimensional structural schematic diagram of the test bed

plane guide way which guides and supports. The schematic diagram of the guide device is shown in Figure 2. Guide rail (7) is fixed in the inside of the pillar by screw(8). There is a copper guide that is fixed beside both sides of the guide, and the copper guide is connected with guidepost(2) by screw. Guidepost is fastened with slider (4) by screw. There is a wedge(3) which is fixed between the slider and the guidepost, and the inclined plane of the wedge is in contact with the inclined plane of the guidepost. By adjusting the contacting inclined plane between guidepost (2) and wedge (3), the precision of the guide is controlled. Gasket (5) is installed in between the slider and the guidepost. The clearance between guidepost (2) and slider 4 can be adjusted by adjusting gasket (5). Set screw (1) is installed in the wedge (3), and the setscrew's upper end passes through wedge (3) to support and act on the slider. By the upward motion of set screw (1), the contacted the guidepost moves up and down or left and right, so as to the clearance between copper guide rail(6)and guide rail(7) is adjusted. The adopted guide can endure bigger torque and load, and avoid the dead-lock and asynchronousness when the cutter header propels [5].

Simplifying the cutter head in the shield tunneling machine, the structural schematic diagram of the cutter head is showed in the figure 3. The cutter head is in a



**Fig. 2.** The schematic diagram of the guide device: (1)set screw; (2)guidepost; (3)wedge; (4)slide; (5)gasket; (6)copper guide rail; (7)guide rail; (8)screw; (9)screw for guidepost; (10)screw for wedge; (11)screw for copper guide rail;

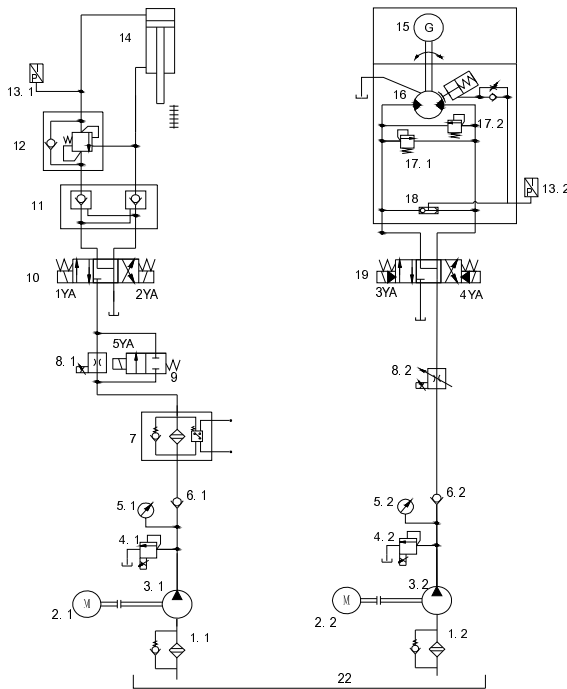


**Fig. 3.** Three-dimensional structural schematic diagram of the cutter header

spoke form, and there are 6 spokes that make up the cutter. The diameter of the cutter head is 1200mm. There are reserved holes in every poke that used to install cutting tools including hobs, cutter, etc. The test bed can be used to single cutter or multi cutter test, and the phase angle and the cutter pitch are adjustable at random. There is a ball type single row slewing bearing installed on the upper surface of the lower beam of the test bed. The above the slewing bearing is a material bin. The slewing bearing's turning motion is driven by a hydraulic motor. The material bin contains the cutting object that is concrete or rock. The diameter of the material bin is 1500mm. The maximal opening height of the cutter header and the material bin is 650mm. The center of the cutter header coincides with the material bin's center.

### 3 Hydraulic Systems of Test Bed

According to the functional requirements of the cutter cutting test bed, the test bed use the manner controlled by hydraulic system. As required real-time adjustment of the load pressure and turning speed, the system is a typical electro-hydraulic proportional control system. Test hydraulic system principle diagram is shown in Figure 4.



**Fig. 4.** Hydraulic system diagram of shield cutter cutting test bed: (1)filter; (2)motor; (3)hydraulic pump; (4)proportional relief valve; (5)pressure gauge; (6)check valve; (7)high pressure filter; (8)proportional velocity regulating valve; (9)two-position two-port electromagnetic valve; (10)three-position four-port electromagnetic valve; (11)hydraulic lock; (12)balance valve; (13)hydraulic pressure transducer; (14) hydraulic cylinder; (15) reducer; (16) hydraulic motor; (17) relief valve; (18) shuttle valve; (19) electro-hydraulic valve;

### 3.1 Cutter Head Propulsion System

The cutting of cutting cutter shield machine test bed can conduct the pushing up and down linear motion along the slide rail promoted by vertical cylinder. The hydraulic system uses quantitative pump, the control part is composed of the proportional relief valve, proportional velocity regulating valve, two-position two-way solenoid directional control valve, three-position four-way solenoid directional control valve, related supporting valve and detecting element. The nominal pressure of oil cylinder is 650KN, and stroke is 700mm. When the cutter header is advancing, the two-position two-port electromagnetic valve (9) loses the power, and the system pressure oil flows through the proportional velocity regulating valve(8.1), and then the left position(1YA) of the three-position four-port electromagnetic valve(10) is electric, and the pressure oil flows through the left position of the valve (10), and then flows into the hydraulic lock(11), and passes through the balance valve (12) and enters the top of hydraulic cylinder (14), and pushes down the hydraulic cylinder piston rod. In the promoting process, the speed of the advance can promote detection tested by build-in displacement transducer in real-time, and change into electric signal and feed back to the proportional velocity regulating valve(8.1), and the opening degree of valve control the promoting speed. The thrust pressure of the hydraulic cylinder is tested by hydraulic pressure transducer (13.1) and changed into electric signal and adjusted into the set pressure of proportional relief valve (4.1), and can adjust propulsion. By adjusting the ratio of proportional control valve and relief valve, the hydraulic cylinder can promote real-time speed and pressure control. Promoting the rollback of hydraulic cylinder, the left position (5YA)of the two-position two-port electromagnetic valve(9) is electric, and the proportional velocity regulating valve (8.1) is shorted, and the system offers large oil, and the right position (2YA)of three-position four-port electromagnetic valve (10) is electric. The pressure oil flows into the lower part of the hydraulic cylinder and pushes the piston rod to roll back quickly. When the working state is stopping, the hydraulic lock(11) can prevent leakage and the hydraulic cylinder does not fall because of gravity. When hydraulic cylinder is being returned lonely, the balance valve(12) can play the role of smooth movements [6-8].

### 3.2 Material Bin Rotary System

The material bin rotary system has some features which are large power and a wide power range. Figure 4 is the main diagram of drive hydraulic system of the rotation material bin. By adjusting the three-position four-way valve electro-hydraulic valve (19) in Fig. 3, it can adjust the rotation direction of hydraulic motor. The speed of rotation material bin is adjusted by the hydraulic motor speed, and the motor speed signal is detected by sensors. After processing, the signal is fed back to the proportional velocity regulating valve(8.2), which regulates valve orifice opening speed control, and can adjust the motor speed. The hydraulic pressure transducer (13.2) tests the pressure of hydraulic motor inlet, and the treated signal feeds back to the proportional relief valve (4.2), and then the system pressure can be adjusted. Relief valves(17.1&17.2) play the role of security [9-10].

## 4 Conclusions

By analyzing the functional requirements of the cutting cutter shield machine test bed, we have designed a test bed, which can fix the cutter header and realize rotation of material bin. A test bed which can fix the cutter header and realize rotation of material bin is developed. The propulsion and loading of the test bed is controlled by the electro-hydraulic proportional control system, the test bed tests the single and multiple factors affect on the cutter header. The hydraulic cylinder working feed, motor rotation speed, direction and working pressure are adjusted continuously in real time, and the material bin can realize the two-way rotation, and three states of hydraulic cylinder working feeds can be realized. The Hydraulic cylinder working feeds has three working states. That is, the constant pressure state, the constant speed state, the constant cutting depth state. The constant pressure state is when the working is feeding, the cutter Hydraulic cylinder working feeds provides the constant pressure as well as the constant thrust force. The constant speed state is the oil cylinder maintains to the constant speed state as well as the constant propulsion speed. The constant cutting depth state is when the working feeds in a certain distance, the cylinder is lockable, and then the motor starts rotating. The designed test bed can accurately and effectively complete the required tasks.

**Acknowledgments.** This research is supported by the Major State Basic Research Development Program of China (973 Program, No. 2007CB714002) and Central South University Postdoctoral Sustentation Fund (No.743410151), which is gratefully acknowledged.

## References

1. Shi, H., Gong, G., Yang, H., et al.: Design and Performance Analysis of Hydraulic Drive System for Cutter head of  $\Phi 3\text{m}$  Shield Prototype. *J. Chinese Hydraulics & Pneumatics* (10), 33–37 (2008)
2. Zhao, X., Hu, G.: Electrical Control System Design of Sllield Simulator Test Rig. *J. Coal Mine Machinery* 29(5), 120–122 (2008)
3. Hu, G., et al.: Simulation and Experimental Analyses of Synchronization Control for Shield Thrust System. *J. China Mechanical Engineering*. 19(10), 1197–1200 (2008)
4. Zhang, Q., Li, S., You, C.: Development and application of new type combination 3d geometrical model test rack apparatus. *J. Chinese Journal of Rock Mechanics and Engineering*. 26(1), 143–147 (2007)
5. Dai, Y.: An inner wedged type four corner and eight plane guide. *Hydraulic Press* (2005), 200420085231.9
6. Hu, G., Gong, G., Yang, H., et al.: Hydraulic System Integration and Experimental Analysis of Shield Tunnelling Machine for Simulator Test Rig. *J. Transactions of The Chinese Society For Agricultural Machinery* 36(12), 102–105 (2005)
7. Chen, K., Han, Y.: Simulation test-bed of slurry shield control system. *J. Construction Mechanization*. 28(4), 59–62 (2007)

8. Zhao, L., Zhou, S., Xie, B.: Remake of Friction Coefficient Bed with Automation. *J. Chinese Hydraulics & Pneumatics* (3), 59–62 (2009)
9. He, Y.: Development of a Test Bench for Control System of Grout Shield. *J. Construction Machinery and Equipment*. 38(11), 28–31 (2007)
10. Hu, G., et al.: Design and Experiment of Hydraulic Propelling System in Shields. *J. Construction Machinery and Equipment*. 38(3), 44–47 (2007)

# The Effect of TBM Hydraulic Piping System Parameters on FSI Vibration\*

Xie Jinghua, Tian Ke, and Yang Dawei

Key Laboratory of Modern Complex Equipment Design and Extreme manufacturing (Central South University), Ministry of Education, Changsha 410083, China

**Abstract.** The characteristics of the hydraulic piping system in Tunnel Boring Machine (TBM) are large slenderness ratio, high fluid pressure and low flow velocity. With the work conditions of TBM, the beam model of long straight pipe was established. Under consideration of fluid structure interaction (FSI) and simple supported conditions, oscillation equation of piping system was solved by vibration formation decomposition. Having analyzed the effect of piping system parameters on the first order inherent frequency by the method of sensitivity analysis, pipe length was the most significant effective factor on inherent frequency of pipe system. Considering the relation between fluid pressure wave frequency and pipe inherent frequency, oscillation criteria based on the first order inherent frequency was put forward. Finally, optimized design was given for piping system design.

**Keywords:** Tunnel Boring Machine, Piping system, fluid structure interaction, beam model, sensitivity analysis.

## 1 Introduction

Tunnel Boring Machine is typical high-power engineering equipment, which includes many complex hydraulic subsystems. Liquid-filled pipelines of the TBM between implementing and power sectors are very long, due to fluid structure interaction, the vibration of pipe is increased, and control signal is lagged feedback. Meanwhile, the effects of FSI are intensified by the valve closure, the changes of loads and even the impurities interfused in the oil. These also result in severe vibration of pipelines, and even there may be extreme conditions such as pipe's cracking. At the same time, the work conditions of TBM are very complex, hydraulic propulsion system is always at variational load conditions as results in real-time changes of soil conditions in pushing process. Thus, it is significant to suppress pipe vibrations, to avoid resonance and to ensure the work safety of TBM's hydraulic system by selecting the parameter of pipeline system [1] [2] [3].

In engineering application of TBM pipeline system, the ratio of piping length to diameter is much larger than 1, as well as the ratio of thickness to diameter isn't too smaller to be ignored, thus beam model is widely be used in the study of pipeline vibration. Paidoussis & Issid proposed a general model to describe lateral motion

---

\* This project is supported by the National Basic Research Program of China (973 Program, Grant No. 2007CB714004).

of long straight pipe [4]. According to this model, many scholars deduced a few of similar form (add or ignore some items), and the conclusive results are: As the fluid velocity increases, natural frequency of the pipeline system is reduced because of the reduction of effective stiffness [5]; tremor instability will not occur at the pipeline which is fixed at both ends; Coriolis force have an important impact on the modality of pipeline system [6]. The purpose of this paper is to present oscillation criteria for the effect of liquid-filled piping system, as well as optimize measures. Based on the beam of long straight pipe, to ignore fluid gravity, structural damping and affect of tension-compression, under consideration of fluid structure interaction (FSI) and simple supported conditions, the process of deduction is described in detail to oscillation equation of piping system, and the front second order inherent frequency are solved. To analyze the first order of inherent frequency by the method of sensitivity analysis, in all of the parameters in the pipe system, length of pipe is the most significant factor to inherent frequency of pipe system. Considered the relation between fluid pressure frequency and pipe inherent frequency, oscillation criteria are described by basing on the first order of inherent frequency; at the end of the paper, combined with piping system status, optimize measures are set in design of pipe system.

## 2 Vibration Equation of Straight Pipe by Simply-Supported at Both Ends

Supposed fluid was stable, non-viscosity and incompressible, ignoring the gravity, structural damping and external pull and pressure of the pipe. Then, the beam model of the straight pipe both ends simply-supported is as shown in following Fig.1, the coordinate  $z$  defines the axial direction, while the coordinate  $x$  and  $y$  defines.

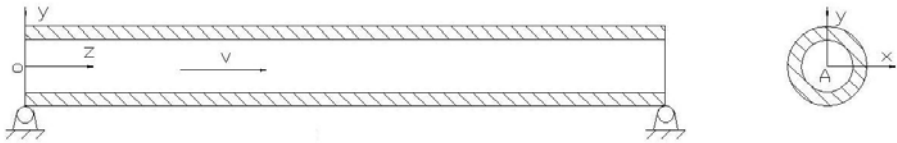


Fig. 1. Diagram of the pipe Coordinate

Adopting the method of Newtonian, describe the continuous pipe system and the equation of motion is as follows [7]:

$$\delta * \int_{t_0}^{t_1} (T - V + W) = 0 \tag{1}$$

In the equation,  $\delta$  is variational operator,  $T$  is the bending momentum of pipes,  $V$  is pipe's deformation potential,  $W$  is the fluid's lateral momentum.

For the straight both ends simply-supported conveying fluid pipe, suppose that the pipeline's and fluid's lateral displacement are independent of each other, but meet the conditions of fluid flowing in the pipe under the conditions of fluid-structure interaction. So then:

$$T = \frac{1}{2} \int_0^l \rho_p A_p \left( \frac{\partial^2 y}{\partial t^2} \right) dz \tag{2}$$



$$V = \frac{1}{2} \int_0^l EI \left( \frac{\partial^2 y}{\partial t^2} \right) dz \tag{3}$$

$$W = \frac{1}{2} \int_{z_1}^{z_2} \rho_f A_f \left( \frac{\partial^2 y(Z,t)}{\partial t^2} \right) dZ \tag{4}$$

Application of Hamilton variation principle, we can gain the interaction vibration equations and boundary conditions as follows [8]:

$$\begin{cases} EI \frac{\partial^4 y}{\partial z^4} + MV^2 \frac{\partial^2 y}{\partial z^2} + 2MV \frac{\partial^2 y}{\partial t \partial z} + (m + M) \frac{\partial^2 y}{\partial t^2} = 0 \\ y_0 = \frac{\partial^2 y}{\partial z^2} = 0, (z = 0, l) \end{cases} \tag{5}$$

In the equation:  $M = \rho_p A_p$ ,  $m = \rho_f A_f$ ,  $M$ ,  $m$  are mass per unit length of pipe and fluid respectively.  $EI$  is pipe’s bending stiffness,  $V$  is steady-state flow rate.

### 3 Solving Vibration Equation

It supposes that radial displacement of pipe is the summation of various vibration modes, hence,  $y(z, t)$  can be written as follow:

$$y(z, t) = \sum_{j=1}^{\infty} y_j(z, t) \tag{6}$$

which:

$$y_j(z, t) = \sum_{n=1,3,5,\dots} a_n \sin\left(\frac{n\pi z}{l}\right) \sin \omega_j t + \sum_{n=2,4,6,\dots} a_n \sin\left(\frac{n\pi z}{l}\right) \cos \omega_j t \tag{7}$$

Substituting (6) into (5), one obtains in matrix form:

$$[K - m\omega_j^2 I] \{a\} = 0 \tag{8}$$

The solution of equation (8) can be written as follow:  $\{a\}^T = (a_1, a_2 \dots a_n)$ , where  $K$  is square matrix of order  $n$ ,  $I$  is unit matrix. If  $\{a\}^T$  have nontrivial solution, thus:

$$[K - m\omega_j^2 I] = 0 \tag{9}$$

If considering the most simple approximate solution, equation (9) only contains two vibration modes, supposing  $a_j=0(j=3, 4, \dots)$ , the relationship of vibration amplitude between the first and the second vibration modes can be written as:

$$a_1 = a_2 \frac{32m_f V \omega_j}{3l \left[ EI \left( \frac{\pi}{l} \right)^4 - m_j V^2 \left( \frac{\pi}{l} \right)^2 - m\omega_j^2 \right]} \tag{10}$$

This relationship indicates: when  $V$  and  $\omega_j$  have a certain matching relation,  $a_j$  is approaching infinite, which is meaning divergence state. The simplest state is: the denominator of equation (10) is zero, as well as  $\omega_j = 0$ , thus, critical flow velocity  $V_c$  is:

$$V_c = \frac{\pi}{l} \sqrt{\frac{EI}{m_f}} \tag{11}$$

Substituting (11) into (10), the frequency equation is available as follow:

$$\left[ EI \left( \frac{\pi}{l} \right)^4 - m_f V^2 \left( \frac{\pi}{l} \right)^2 - m \omega_j^2 \right] \left[ EI \left( \frac{2\pi}{l} \right)^4 - m_f V^2 \left( \frac{2\pi}{l} \right)^2 - m \omega_j^2 \right] - \frac{256}{9l^2} (m_f V \omega_j)^2 = 0 \tag{12}$$

The first and the second orders inherent frequency  $\omega_j = 0$  ( $j = 0, 1$ ) are:

$$\omega_j^2 = \omega_n^2 \left( A \pm \left\{ A^2 - 4 \left[ 1 - \left( \frac{V}{V_c} \right)^2 \right] \left[ 4 - \left( \frac{V}{V_c} \right)^2 \right] \right\}^{\frac{1}{2}} \right) \quad j = 1, 2 \tag{13}$$

In expression (13):

$$A = 8.5 - \left( \frac{V}{V_c} \right)^2 \left[ 2.5 + \frac{128}{9\pi^2} \left( \frac{m_f}{m} \right) \right] \tag{14}$$

As well as  $\omega_n$  is basic vibration mode of inherent frequency for the pipe of two simple supported ends with still fluid. While the number of simple supported is  $N$ , inherent frequency of long straight pipe with equal spans is

$$f_n = 35.5 \lambda_n \sqrt{\frac{E(d_o^4 - d_i^4)}{ml^4}} \tag{15}$$

Where  $\lambda_n$  represent frequency constant, thus,

$$\omega_n = 2\pi f_n = 71\pi \lambda_n \sqrt{\frac{E(d_o^4 - d_i^4)}{ml^4}} \tag{16}$$

### 4 Sensitivity Analysis of the First Order Inherent Frequency

Take the TBM of the Guangzhou Metro Line 2 as the model, the parameters were selected as following: flow velocity was  $V_f = 1m/s$ , the thickness of pipe was  $\delta = 0.003m$ , the elastic modulus of the pipe was  $E = 210GPa$ , the mass density was  $\rho_p = 7800kg/m^3$ , fluid density was  $\rho_f = 890kg/m^3$ , the bulk modulus was  $K = 2.1Gpa$ . Single-span of pipe was set between  $3m$  and  $6m$ , considered one pipe clip would be loose in actual work, the length of pipe was selected between  $3m$  and  $15m$  in sensitivity analysis. The radius of pipe was set between  $0.01m$  and  $0.07m$ .

### 4.1 The Sensitivity of the Pipe Length to First Order Inherent Frequency

In Fig.02, it shows the sensitivity of the pipe length to first order inherent frequency when  $R=0.01m$  (red solid line),  $R=0.03m$  (blue broken line),  $R=0.05m$  (green solid-dot line),  $R=0.07m$  (black dot line). The sensitivity is negative, which means as the length of pipe increasing, first order inherent frequency is decreased. When the length is greater than  $9m$ , the sensitivity keeps constant, close to  $0$ . This situation should be avoided in the pipeline design process.

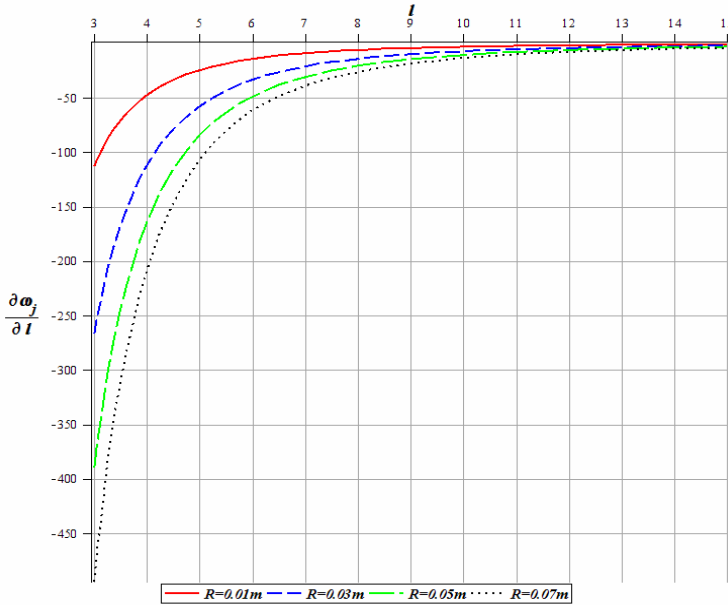


Fig. 2. The sensitivity of the pipe length to first order inherent frequency

The length of pipe is between 3m to 8m, the numerical value of the sensitivity is present in table01:

Table 1. The sensitivity numerical value when pipe length between 3m to 8m

The radius of pipe(m)	The length of pipe (m)					
	3	4	5	6	7	8
0.01	-112.73	-47.56	-24.35	-14.09	-8.87	-5.94
0.03	-266.38	-112.38	-57.54	-33.3	-20.97	-14.05
0.05	-389.86	-164.47	-84.21	-48.73	-30.69	-20.56
0.07	-494.49	-208.61	-106.81	-61.81	-38.92	-26.08

Supposed threshold value was -10, based on the parameter of TBM pipeline system, the range of pipe radius range from 0.03m to 0.05m. According to table01, the length of single span pipe can be set from 3m to 5m. Due to this, if there is one pipe clip loosened, it will be still able to satisfy the safe requirements of the TBM working conditions.

### 4.2 The Sensitivity of the Pipe Radius to First Order Inherent Frequency

Fig.03 shows the sensitivity of the pipe radius to first order inherent frequency when  $L=6\text{ m}$  (red dot line),  $L=9\text{ m}$  (blue solid-dot line),  $L=12\text{ m}$  (green broken line),  $L=15\text{ m}$  (black solid line).Mainly, when the pipe length is a fixed value, the effect of the pipe radius to first order inherent frequency is reduced by the pipe radius increasing. When pipe length is longer than 9m, the effect of pipe radius to inherent frequency tends to the similar. As the pipe length is between 3m to 9m, the effect of pipe radius to inherent frequency is greater by the pipe radius reducing. In Fig03, when pipe radius is between  $0.01\text{m}$  and  $0.06\text{m}$ , the sensitivity shows the following characteristics  $\frac{\partial \omega_1}{\partial R}(L=6) > \frac{\partial \omega_1}{\partial R}(L=9) > \frac{\partial \omega_1}{\partial R}(L=12)$ ; when the pipe length is beyond 0.06m, the sensitivity keeps one constant.

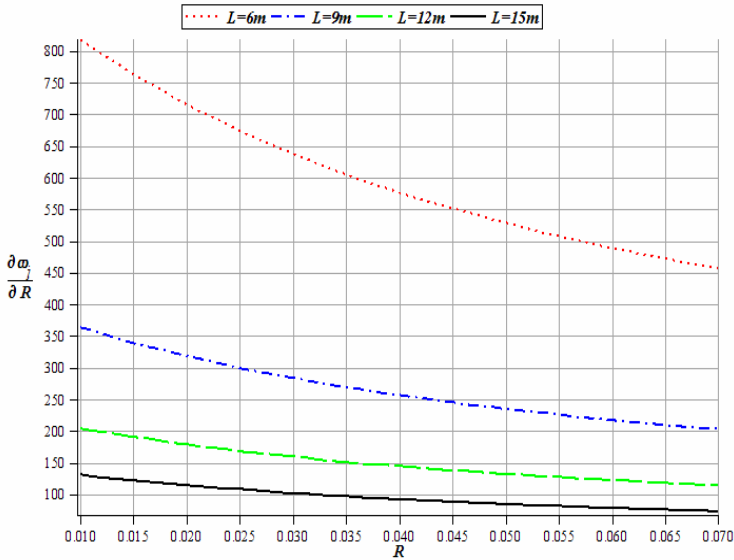
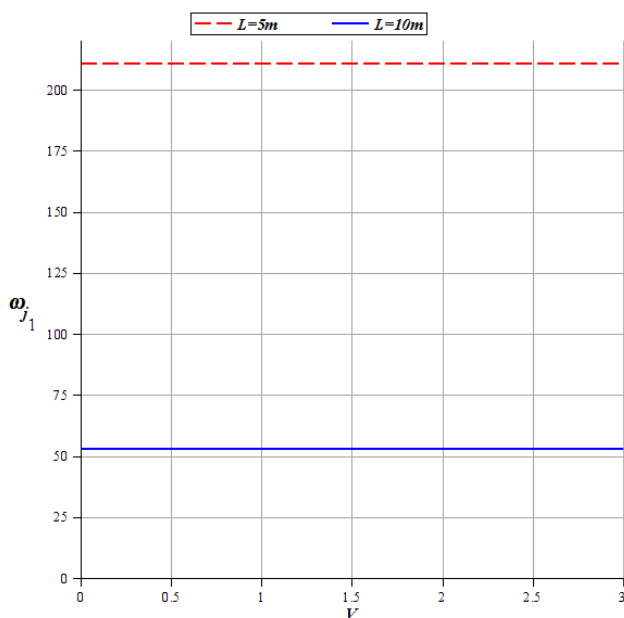


Fig. 3. The sensitivity of the pipe radius to first order inherent frequency

### 4.3 The Effect of Flow Velocity to the Pipe Length to First Order Inherent Frequency

The pipe length of single span is selected as 5m, the pipe radius is 0.05, other parameter is as the same as above, the effect of flow velocity to the pipe length to first order inherent frequency is shown in Fig.04.



**Fig. 4.** The effect of flow velocity to the pipe length to first order inherent frequency

Because the flow velocity in TBM pipeline system is low, supposed the range of the flow velocity was from 0m/s (quiescent condition) to 3m/s, the first order inherent frequency appears no changes in Fig.04. It proves the low flow velocity have no significant effect in inherent frequency of pipe.

## 5 Oscillation Criterion of Long-Distance Pipeline

When the length of the pipe reaches threshold value, resonance phenomenon is easy to occur between the pipe and fluid. Only when pipe frequency is less than a certain value of fluid frequency, the occurrence of resonance phenomenon will be avoided, and the effect of vibration on the pipe and fluid will be reduced. For the liquid-filled piping system, frequency of oil pressure contains two types: coupling frequency of oil pumping from hydraulic pump, and frequency of the vibration caused by loading effect of valves in piping system. Moreover, the latter is the major factor.

Frequency of oil pressure is normally written as follow:

$$\omega = \frac{\pi a_f}{2l} \quad (17)$$

When the ratio of excitation frequency and pipe inherent frequency is greater than 0.5, probability of vibration is very large. As written as following:

$$\frac{\omega}{\omega_j} \geq C \quad (18)$$

The constant  $C$  is decided together by the pipeline damping and the half-power bandwidth of resonance region. According to literature [9], it is assumed the constant  $C$  equal to 0.5. Substituting (13)(17) into (10), one obtains:

$$l \leq \frac{71\lambda_n}{a_f} \sqrt{\frac{E(d_o^4 - d_i^4)}{m}} \sqrt{A - \{A^2 - 4[1 - (\frac{v}{v_c})^2][4 - (\frac{v}{v_c})^2]\}^{1/2}} \quad (19)$$

Note that equation (19) is set, it can be deemed the vibration of pipe is within bounds; if equation (19) is NOT set, the vibration is visible and need to impose damping measures.

While limiting velocity  $V_c$  is very large and the practical fluid velocity  $V_f$  is limited, it approximately is  $V_f/V_c=0$ . Equation (19) can be simplified as:

$$l \leq \frac{400\lambda_n}{a_f} \sqrt{\frac{E(d_o^4 - d_i^4)}{m}} \quad (20)$$

## 6 Conclusion

- (1) In the case of the selected radius ( $R=0.03-0.05m$ ), the length of single-span pipeline should be limited to the range of  $3-5m$ . Due to this, even if there was one pipe clip loosed, it would be able to satisfy the safe requirements of TBM working conditions. Meanwhile, both properly increasing the diameter and wall thickness and using the pipe with high elastic modulus certainly make contributions to pipeline system safety.
- (2) Considering the relation between fluid pressure wave frequency and pipe inherent frequency, oscillation criteria based on the first order inherent frequency was put forward. It can be simplified as Equation (20).
- (3) Building the model of piping system, tube is not involved in this paper. Future research will focus on modeling the complex piping system to provide design criteria for piping system.

## References

1. Zhang, L.X., Yang, K.: Fluid-structure interaction theory and its application, pp. 34–38. Science and Technology Press, Beijing (2004) (in Chinese)
2. Chen Y-s.: With regard to non-linear vibration. J. Advances In Mechanics 2, 59–66 (1984)
3. Zhang, L.X., Tijsseling, A.S., Vardy, A.E.: FSI analysis of liquid-filled pipes. J. Sound and Vibration. 224, 86–92 (1999)
4. Paidoussis, M.P., Issid, N.T.: Dynamics stability of pipe conveying fluid. J. Journal of Sound and Vibration 33(2), 267–294 (1974)
5. Benjamin, T.B.: Dynamics of a system of articulated pipes conveying fluid. Proceeding of the Royal Society A261, 457–499 (1961)
6. Naguleswaran, S., Williams, J.H.: Lateral vibration s of a pipe conveying fluid. J. Journal of Mechanical Engineering Science 10, 228–238 (1968)

7. Meirovitch, L.: Analytical Methods in Vibrations. The Macmillan Co. Ltd., New York (1967)
8. Paidoussis, M.P.: Fluid-Structure Interactions. Slender Structures and Axial Flow 1, 16–18 (1998)
9. Dayou, M.: Noise and vibration control engineering, pp. 32–37. China Machine Press (CMP) (2002)

# Station Point Optimization of Total Station in Shield Tunneling Method Based on Set-Covering

Bolan Li, Xu Zhang, and Limin Zhu\*

State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China  
zhulm@sjtu.edu.cn

**Abstract.** This paper proposed a method for station point optimization of the total station of automatic guidance system in shield tunneling, which is not analyzed widely at present. In this paper, station point optimization is converted into a set-covering problem. In order to express the measurability of the active target station for the total station, a 0-1 measurability matrix is introduced, and each element in the matrix represents a measurability judgment, which includes the visibility test, the distance test, and the incidence angle test. Then, the greedy algorithm is used to find the approximate optimal solution of this set-covering problem. Furthermore, an application example affirmed that our method was effective.

**Keywords:** total station, station point optimization, set-covering, shield tunneling.

## 1 Introduction

Nowadays, the rapid development of world economy accelerates the progress of urbanization. Along with the enhancement of urban density and the emergence of more and more skyscrapers, available space on the ground shrinks, so that how to efficiently utilize underground space becomes an important task in urbanization, and shield tunneling method is the best choice at present [1]. In China, the demand for tunnel construction is great.

During the past few years, the creation of subterranean space by means of shield tunneling became increasingly established. These spaces had mainly been used for dealing with traffic. The general principle of the shield is based on a cylindrical steel assembly pushed forward on the axis of the tunnel while at the same time excavating the soil [2].

In the classic type of guidance system in shield tunneling, geodetic engineers played a key role to give steering information for the construction work, and it was hard to achieve a real-time measurement. Every time when geodetic engineers needed to measure the position and orientation of the tunnel boring machine (TBM), the

---

\* Corresponding author.



construction had to stop for the measurement. Due to modern geodetic sensors and processing techniques, a highly precise automatic guidance system [3] was developed to control the TBM to excavate along designed tunnel axis (DTA).

This automatic system primarily consists of the total station and the active target station (ELS). Normally, the total station is set on the tunnel wall, and the ELS is attached to the rear of the TBM. The total station emits laser beam aiming at the ELS to survey the position and orientation of the TBM. The premise of this system is that the ELS must be in line of sight of the total station. With the TBM excavating forward, the total station has to be transferred to the next station point when the ELS gets out of its field of view. The transfer of the total station means that a new station point in the tunnel ahead should be selected for setting the total station to ensure continuous measurement. But, the transfer process of the total station is a time-consuming task, and it includes many steps as follows: building support lug at the new station point, setting survey prisms, surveying positions of prisms, disassembling the total station at the old station point, assembling the total station at the new station point, and checking the coordinate of the new station point to generate the new geodesic reference frame. The process should generally be applied for several times to get precise steering information, and at the same time the construction has to stop, which inevitably lower the efficiency of the whole construction. More transfer times mean longer construction time and higher cumulative error, so reducing transfer times helps to shorten construction time, improve construction quality, and further, cut down the construction cost. Unfortunately, few literatures or publications that focus on the research of this process have been found, and the principle of the transfer, where the next station point is located, is still under the guidance of experience. Undoubtedly, for geodetic engineers, the best situation is to realize fewest station points to fill measurement requirements of the construction. In this paper, a method to optimize station points for the total station based on set-covering problem (SCP) is proposed. Provided DTA and the model of both the TBM and its accessory components, station points can be optimized in advance to the construction phase, and that would offer a guideline for geodetic engineers.

The structure of this paper is as follows. In Section 2, according to actual construction process, the model of the transfer of the total station is established, and converted into a set-covering problem. After that, the outline of the approach is given. Section 3 presents an application example with some analysis. Section 4 draws a conclusion and the future work is discussed.

## 2 The Mathematical Model for the Transfer

Fig. 1 shows the relative position between the total station and the ELS. Here, point  $a$  and point  $b$  represent positions of the total station and the ELS, and meanwhile  $o_a$  or  $o_b$  on the DTA are called the basis point. Positions of the total station and the ELS can be determined by  $o_a$  and  $o_b$ . From each  $o_a$  can generate a circle to represent the tunnel wall at this point for setting the total station, while each  $o_b$  corresponds to a unique position for the ELS.

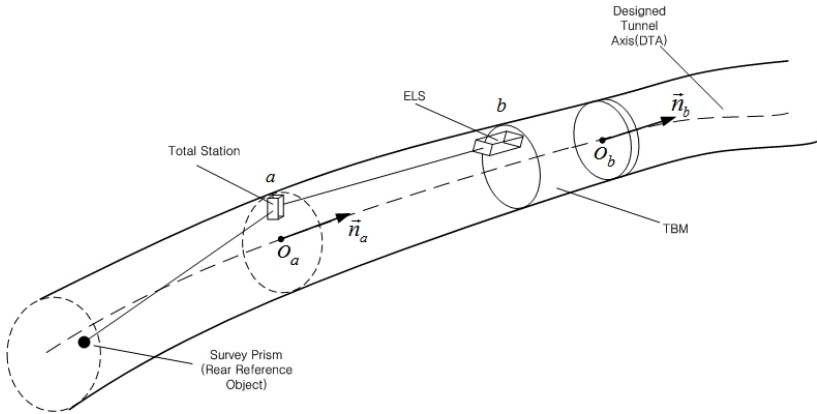


Fig. 1. Automatic guidance system for shield tunneling

### 2.1 Constraints in Automatic Guidance System

In shield tunneling, the complex and variable environment means many factors influence the measurability between the total station and the ELS, and these factors can be classified into three categories:

**Visibility Constraint.** Accessory components of the TBM, such as erector, screw conveyor, and trolley for carrying tunnel segments, can easily block the laser beam path from the total station to the ELS. Next, DTA is a complicated spatial curve so that the ELS would be out of the visibility field of the total station when the TBM moves along such DTA, because the tunnel wall may block the ray. Furthermore, the orientation of the TBM is continually changing during excavation, which can not make sure that the total station can always trace the ELS. Given DTA and the CAD model of the TBM, these problems are equivalent to a visibility test in computer graphics. The test function is defined as:

$$V(a,b) = 0 \text{ or } 1 \tag{1}$$

where  $V(a,b)=1$ , if the ELS is visible for the total station,  $V(a,b)=0$  otherwise.

**Distance Constraint.** The air environment in the tunnel is not at a homogeneous and stable state, which can weaken the transmission of laser beam. Then, as an electronic equipment, the total station has its measurement range. Therefore, the distance  $l$  from the total station to the ELS must be less than  $l_{max}$ . Besides, the space for the TBM and its accessories requires that  $l$  be greater than  $l_{min}$ . The constraint can be described as:

$$l_{min} \leq l \leq l_{max} \tag{2}$$

**Incidence Angle Constraint.** For the ELS, the angle  $\theta$  between the incidence ray and its normal direction should be less than  $\theta_{max}$  to guarantee the measurement precision. Let  $\vec{ab}$  be the unit incidence ray and  $\vec{n}$  represent the normal direction of the ELS. The incidence angle constraint can be written as:

$$\vec{ab} \cdot \vec{n} \geq \cos(\theta_{max}) \tag{3}$$

### 2.2 The Set-Covering Model

**Definition 1 (The position space of the total station).** As shown in Fig. 1, the position  $a$  of the total station is on the circle whose center is  $o_a$ . The position space of the total station (TS) is defined as:

$$C = \{a \mid |o_a a| = R \text{ and } \overrightarrow{o_a a} \cdot \vec{n}_a = 0, o_a \in D\} \tag{4}$$

where  $D$  denotes the space of DTA,  $\vec{n}_a$  is the tangent direction of point  $o_a$ , and  $R$  is the radius of the tunnel.

**Definition 2 (The position space of the ELS).** The position  $b$  of the ELS is given under the local reference frame of the TBM. The position space of the ELS (ES) is defined as:

$$U = \{b \mid b = F(\omega_b, o_b, \vec{n}_b), o_b \in D\} \tag{5}$$

where  $\omega_b$  is the local coordinate of the ELS in the TBM reference frame, and  $\vec{n}_b$  is the tangent direction of point  $o_b$ .

For each point  $a \in C$ , through the visibility test, the distance test and the incidence angle test, a subset  $u \subset U$  can be obtained. It means that the point  $a$  can see any point  $b \in u$ , and  $u$  is a subset of  $U$ . If we select a point set of  $\{a_i\}$  from  $C$ , we can get a set  $S = \{u_i \mid i = 1, 2, \dots\}$ , where  $u_i \subset U$ . The purpose is to find the smallest  $T \subset S$  that covers all points of  $U$ :

$$\bigcup_{u_i \in T} u_i = U \tag{6}$$

### 2.3 The Measurability Matrix

The set-covering model in Section 2.2 cannot directly be applied for computer calculation. Actually, the set-covering model can be converted into a 0-1 matrix [4]. In this section, a 0-1 measurability matrix is established. Here, we sample a set of  $\{a_i\}$  from  $C$  and a set of  $\{b_j\}$  from  $U$  to represent  $C$  and  $U$ , respectively, where  $i, j = 1, 2, \dots$ .

**Definition 3 (Constraint matrixes for constraint tests between the total station and the ELS).** According to three constraints (1), (2) and (3) analyzed in Section 2.1, the visibility constraint matrix (VCM), the distance constraint matrix (DCM) and the incidence angle constraint matrix (ICM) can be introduced to describe these three constraints. They are defined as follows:

$$\text{VCM} \rightarrow V(a_i, b_j) = \begin{cases} 1 & \text{visible} \\ 0 & \text{invisible} \end{cases} \tag{7}$$

$$\text{DCM} \rightarrow D(a_i, b_j) = \begin{cases} 1 & l_{\min} \leq |a_i b_j| \leq l_{\max} \\ 0 & |a_i b_j| > l_{\max} \text{ or } |a_i b_j| < l_{\min} \end{cases} \tag{8}$$

$$ICM \rightarrow I(a_i, b_j) = \begin{cases} 1 & \overline{a_i b_j} \cdot \overline{n_j} \geq \cos(\theta_{\max}) \\ 0 & \overline{a_i b_j} \cdot \overline{n_j} < \cos(\theta_{\max}) \end{cases} \quad (9)$$

where  $i, j = 1, 2, \dots$ .

**Definition 4 (Measurability matrix for the total station and the ELS).** From (7), (8) and (9) in Definition 3 can generate a measurability matrix  $E$  through logical AND operation to depict the measurability relationship between the total station and the ELS in the whole shield tunneling. The matrix is defined as:

$$E(a_i, b_j) = \begin{matrix} & b_j & \rightarrow \\ a_i & \begin{pmatrix} 1 & 0 & 0 & 1 & \dots \\ 0 & 0 & 1 & 1 & \dots \\ 1 & 1 & 0 & 0 & \dots \\ 1 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} & \\ \downarrow & & \end{matrix} \quad i = 1, 2, \dots, \quad j = 1, 2, \dots \quad (10)$$

where  $E(a_i, b_j) = V(a_i, b_j) \& D(a_i, b_j) \& I(a_i, b_j)$ , the horizontal axis indicates possible positions of the ELS, and the vertical axis represents candidate station points of the total station for optimization.

Then, the set-covering problem is converted into a mathematical matrix model. The measurability matrix  $E = (e_{ij})$  is a 0-1  $m \times n$  matrix, and the  $i$ th row vector  $a_i = (1 \ 0 \ 0 \ 1 \ \dots)_{1 \times n}$  in this matrix gives the measurability judgment results when the total station is set at point  $a_i$ . The  $j$ th element in  $a_i$  represents whether the ELS at position  $b_j$  is measurable for the total station at  $a_i$ . Let  $c = (c_i)$  be a  $m$ -dimensional integer vector,  $M = \{1, 2, \dots, m\}$  and  $N = \{1, 2, \dots, n\}$ . The value  $c_i (i \in M)$  represents the cost of row  $i$ , and it assumes without loss of generality  $c_i > 0$  for  $i \in M$ . It is said that a row  $i \in M$  covers a column  $j \in N$  if  $e_{ij} = 1$ . SCP calls for a minimum-cost subset  $\delta \subseteq M$  of rows, such that each column  $j \in N$  is covered by at least one row  $i \in \delta$ . The mathematical model for SCP is

$$v(\text{SCP}) = \min \sum_{i \in M} c_i x_i \quad (11)$$

subject to

$$\sum_{i \in M} e_{ij} x_i \geq 1, \quad \forall j \in N, \quad (12)$$

$$x_i \in \{0, 1\}, \quad i \in M, \quad (13)$$

where  $x_i = 1$  if  $i \in \delta$ ,  $x_i = 0$  otherwise.

Set-covering problem is a well-known NP-hard optimization problem, see Garey and Johnson [5], which usually cannot be solved exactly within a reasonable amount of time. One of the best polynomial time algorithms for approximating set-covering is the greedy algorithm [6]: at each step choose the unused set which covers the largest number of remaining elements.

## 2.4 Outline of the Approach

Fig. 2 shows the process to optimize station points for the total station of automatic guidance system in shield tunneling. In order to make the visibility test, the TBM and its accessories have to be tessellated into a group of triangles, and they can be generated by commercial CAD software. Meanwhile, a set of  $\{a_i\}$  and a set of  $\{b_j\}$  should be sampled from  $C$  and  $U$  by appropriate intervals for computer calculation, where  $i = 1, 2, \dots, m$  and  $j = 1, 2, \dots, n$ . In this paper, the visibility test is calculated based on Graphic Processing Unit (GPU) with its depth buffer and collision query functionality [7].

The proposed approach in this paper consists of three steps: data input, measurability matrix establishment, and station point optimization. For data input, it includes DTA and the CAD model of the TBM. Then, three tests are implemented to generate the VCM, DCM, and ICM to establish the measurability matrix. The measurability matrix is obtained through logical AND operation of such three matrixes. Finally, after the establishment of the measurability matrix, station point optimization is converted into a set-covering problem, and it is solved by greedy algorithm. In our experiment, approximate optimal station points are obtained.

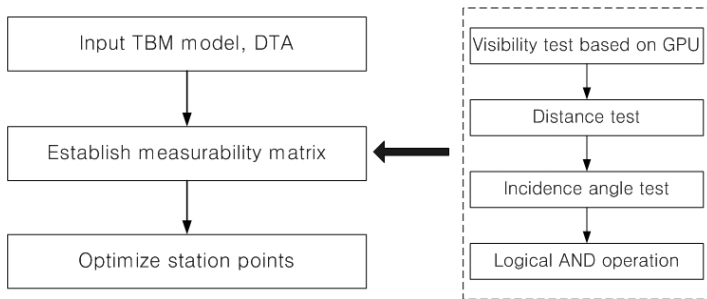


Fig. 2. The flow chart of station point optimization

## 3 Experiment

The presented approach is implemented in an actual example. In this instance, the diameter of the TBM is  $6.28m$ , the length of DTA is  $1080m$  and the local coordinate of the ELS in the TBM reference frame is  $(-5.494m, 0m, 2.190m)$ . The distance constraint is  $10m \sim 200m$  and the incidence angle constraint is  $45$  degree. The smallest radiuses of the horizontal alignment and the vertical alignment of DTA are  $400m$  and  $2990m$ , respectively. The DTA is discretized into 108 line segments by  $10m$ , corresponding to

108 basis points on the DTA. The proportional errors of the horizontal alignment and the vertical alignment can be calculated as follows:

$$\begin{cases} e_{hor} = (R_{hor} - \sqrt{R_{hor}^2 - (l/2)^2}) / R_{hor} = 0.00781\% \\ e_{ver} = (R_{ver} - \sqrt{R_{ver}^2 - (l/2)^2}) / R_{ver} = 0.000140\% \end{cases} \quad (14)$$

where  $R_{hor}$  and  $R_{ver}$  are the smallest radius of the horizontal alignment and the vertical alignment, respectively,  $l$  is the length of one unit line segment. The circle of each basis point is discretized to 360 points by 1 degree step, and that is to say there are 360 candidate station points of the total station at this basis point. Fig. 3 depicts the tunnel model.

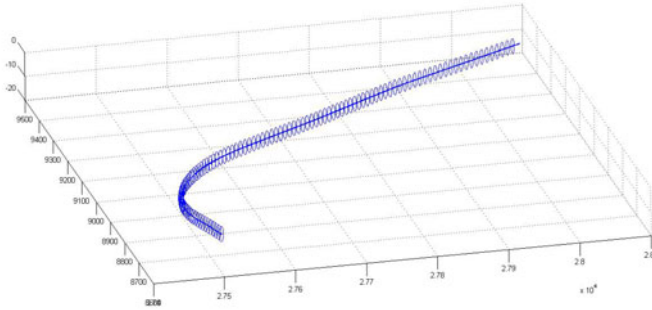


Fig. 3. The model of the tunnel

### 3.1 The Visibility Cone

In order to reduce computing time, a visibility cone (VC) is introduced for the visibility test, which is showed in Fig. 4. The visibility cone is defined as [8]:

$$VC(b, \mathbb{R}) = \{\vec{v} \mid (b - \lambda\vec{v}) \cap \mathbb{R} = \Phi, \forall \lambda > 0\} \quad (15)$$

where  $\mathbb{R}$  represents spatial obstacles between the total station and the ELS, and  $\vec{v}$  is the candidate orientation that can be generated by tessellating Gaussian sphere [9]. In this paper, VC is rewritten as  $VC(b, \vec{t}_1 \dots \vec{t}_r)$  in which  $\vec{t}_1 \dots \vec{t}_r$  mean boundary vectors of VC, so a visibility polygon  $T_1 T_2 \dots T_r$  can be obtained when  $VC(b, \vec{t}_1 \dots \vec{t}_r)$  is projected to the circle plane of one basis point. Then, we can detect whether each of 360 points around the circle is in the polygon  $T_1 T_2 \dots T_r$ . If we get some point  $a$  in the polygon, we have to judge whether line  $ab$  is in the tunnel, and the method is to make sure that the distance between DTA and any point on line  $ab$  should be less than  $R$ . Therefore, if a point  $a$  is both in VC and the tunnel,  $a$  is visible for  $b$ . Fig. 5 shows the calculation result of VC in this example, and all boundary vectors are calculated in the TBM reference frame. This visibility cone can be computed with a GPU-based method. The GPU-based method means that the visibility test is implemented by using collision query functionality of a graphics card.

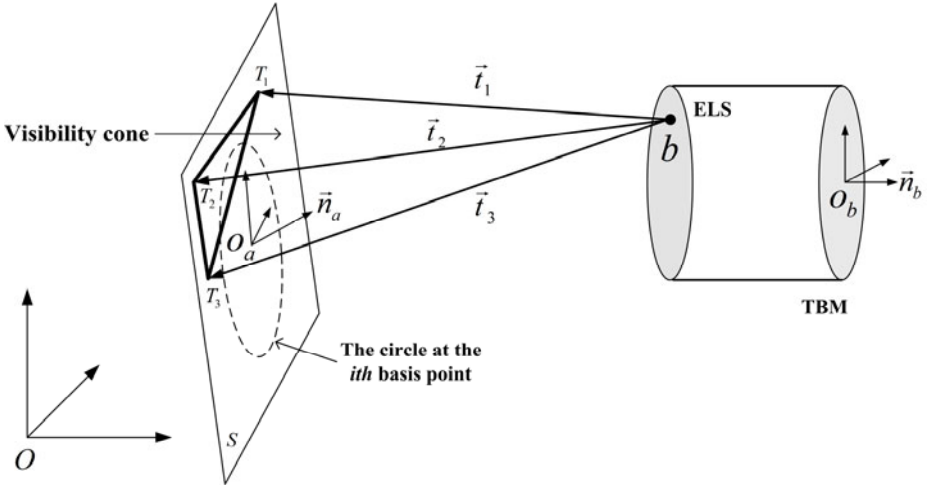


Fig. 4. The visibility cone between the total station and the ELS

$$\begin{aligned}
 \vec{t}_1 &= [-0.875539, 0.407260, -0.259945] \\
 \vec{t}_2 &= [-0.896809, 0.358027, -0.259906] \\
 \vec{t}_3 &= [-0.911772, 0.354576, -0.207239] \\
 \vec{t}_4 &= [-0.923576, 0.350517, -0.155385] \\
 \vec{t}_5 &= [-0.941956, 0.298907, -0.152887] \\
 &\bullet \\
 &\bullet \\
 &\bullet \\
 \vec{t}_{72} &= [0.838575, 0.409079, -0.359788] \\
 \vec{t}_{73} &= [-0.858274, 0.408749, -0.310307]
 \end{aligned}$$

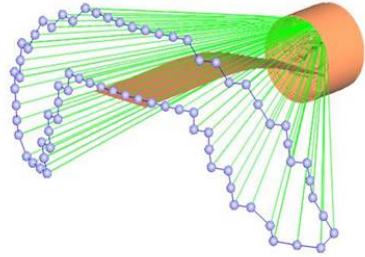


Fig. 5. The calculation result of the visibility cone

### 3.2 The Measurability Matrix

Under three constraints, the visibility constraint, the distance constraint and the incidence angle constraint, the measurability matrix is established as follows:

$$E(a_i, b_j) = \begin{matrix} & b_j & \rightarrow \\ a_i & \downarrow & \\ & \begin{pmatrix} 1 & 1 & 1 & 1 & \dots \\ 1 & 1 & 1 & 1 & \dots \\ 1 & 1 & 1 & 1 & \dots \\ 1 & 1 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} & \end{matrix} \quad (16)$$

$38880 \times 108$

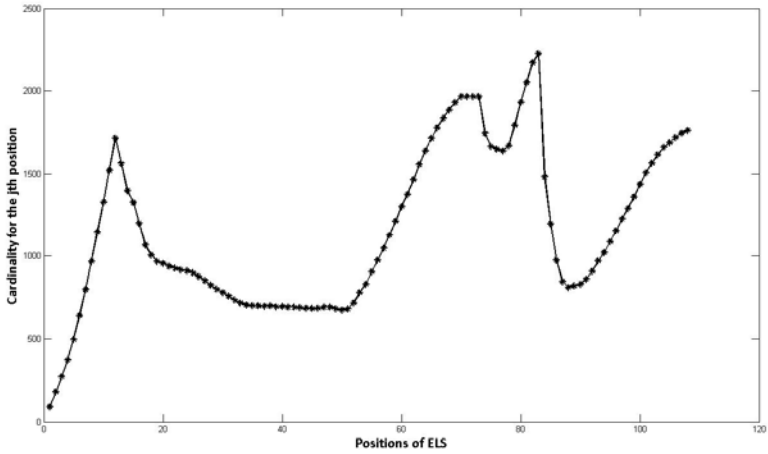
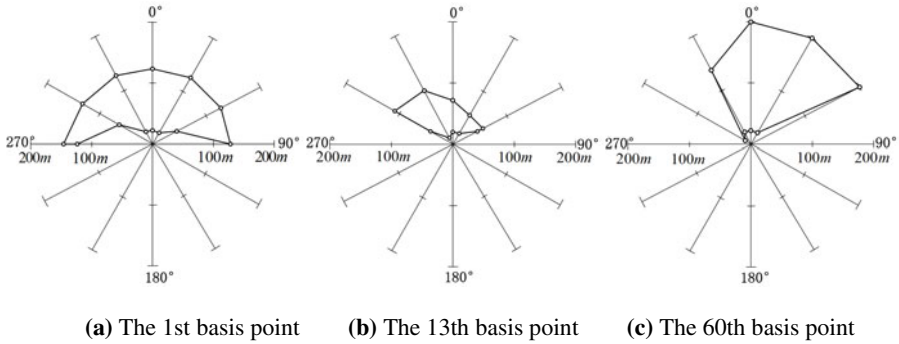


Fig. 6. The variation of the cardinality of DTA



(a) The 1st basis point      (b) The 13th basis point      (c) The 60th basis point

Fig. 7. The tendency of the measurability range

Due to discrete 360 points around each circle, the number of candidate station points for the total station is  $360 \times 108 = 38880$ , corresponding to 38880 rows in the matrix. Similarly, 108 possible positions for the TBM/the ELS mean 108 columns in the matrix. For the  $j$ th position of the TBM/the ELS, there are several candidate station points of the total station that can see the ELS. Here, the amount of such candidate station points is called the cardinality of the  $j$ th position. The sum of elements in the  $j$ th column is the cardinality, and Fig. 6 indicates the variation of cardinalities along DTA. The results show that the higher the cardinality is, the smoother DTA will be.

The variation of measurability range is plotted in Fig. 7 in which (a), (b) and (c) show measurability ranges at the 1st, 13th and 60th basis points, respectively. View orientation from the total station to the ELS is inward perpendicular to paper screen, and the axis depicts the measurability range from 0m to 200m for each point around the circle. The region surrounded by bold lines is the measurability range at each basis point, and the variation of the region corresponds to the tendency of DTA in Fig. 3. For

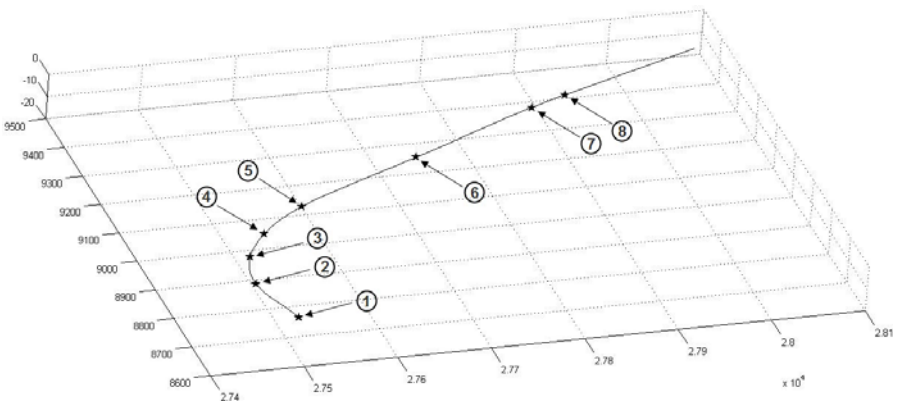


**Table 1.** The optimized results

Station Point	Basis Point	Degree	Row in the Matrix
1	1	317°	317
2	15	306°	5346
3	26	311°	9311
4	34	313°	12193
5	44	314°	15794
6	64	9°	22689
7	84	319°	30199
8	89	1°	31681

example, the 13th basis point is on a high curvature line that extends toward the lower right direction, which leads to such a small region in Fig. 7(b). According to analysis above, it can be concluded that this measurability matrix can precisely depict the measurability between the total station and the ELS.

Based on greedy algorithm, the approximate optimal results are obtained in Table 1, and relative positions of these points can be seen in Fig. 8. In actual construction, the experience guideline of the transfer of the total station is that it is about 100m in the curve segment and 200m in the straight line segment. The optimized results coincide with the experience guideline well.

**Fig. 8.** The optimized station points along DTA

## 4 Conclusion and Further Work

In summary, a matrix approach based on set-covering is presented to optimize station points for the total station in automatic guidance system. Three steps are data pre-process of both DTA and the TBM model, measurability matrix establishment for

set-covering problem, and algorithm to optimize station points. The spatial obstacles have to be tessellated into triangles, preparing for the visibility test. Within permitted error, appropriate intervals should be selected in sampling TS, ES and DTA for computer calculation. The measurability matrix is generated from three tests, through which station point optimization can be converted into a set-covering problem. The visibility test is implemented by a visibility cone achieved through GPU collision query functionality. Greedy algorithm is finally applied to solve the set-covering problem. An actual example and some analysis verify the validity of the approach.

Unlike the ideal consideration in the approach discussed above, there are deviations in pieced tunnel and the position and orientation of the TBM, and these deviations will be studied in the future research.

## Acknowledgments

This work was partially supported by the National Natural Science Foundation of China under grants No.50821003, the National Key Basic Research Program under grant No.2007CB714005, and the Science & Technology Commission of Shanghai Municipality under grant No.09520701700.

## References

1. Chow, B.: *Shield Tunneling Technology*. China Architecture and Building Press, Beijing (2004)
2. Maidl, B., Herrenknecht, M., Anheuser, L.: *Mechanised Shield Tunneling*. Ernst and Sohn, Berlin (1995)
3. Niemeier, W.: *Geodetic Techniques for the Navigation, Guidance and Control of Construction Processes*. In: 3rd IAG/12th FIG Symposium, Baden (2006)
4. Caprara, A., Toth, P.: Algorithms for the Set Covering Problem. *Annals of Operations Research* 98, 353–371 (2000)
5. Garey, M.R., Johnson, D.S.: *Computers and Intractability: A guide to the theory of NP-Completeness*. W. H. Freeman, San Francisco (1979)
6. Slavik, P.: A Tight Analysis of the Greedy Algorithm for Set Cover. In: *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, pp. 435–441. ACM Press, New York (1996)
7. Dave, S.: *OpenGL Programming Guidance*, 5th edn. Addison Wesley Professional, Boston (2005)
8. Bi, Q., Wang, Y., Ding, H.: A GPU-based Algorithm for Generating Collision-free and Orientation-smooth Five-axis Finishing Tool Paths of a Ball-end Cutter. *International Journal of Production Research* 48, 1105–1124 (2010)
9. Dhaliwal, S., Gupta, S.K., Huang, J., Priyadarshi, A.: Algorithm for Computing Global Accessibility Cones. *Journal of Computing and Information Science in Engineering* 3(3), 653–663 (2003)

# Study on Earth Pressure Acting on Cutter Face Based on Nonlinear Optimization

Xiangtao Hu, Yong'an Huang, and Zhouping Yin

State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science & Technology, 430074, Wuhan, P.R. China  
hust\_hoo@163.com

**Abstract.** Shield tunneling technologies have been developed for constructing tunnels in soft ground. It is of major importance to study the earth pressure acting on cutter face. Nowadays, the methods of studying earth pressure could be classified into two categories: Rankine theory and limit equilibrium method. However, these methods cannot represent the distribution of earth pressure on cutter face and dynamic characters. In this paper, Duncan-Chang nonlinear elasticity model is applied to calculate the earth stiffness, and Peck's formula is used to establish the relationship between the ground settlement and earth pressure. Furthermore, an earth pressure model acting on cutter face is proposed based on the principle of minimum potential energy. The proposed model not only can provide the distribution of earth pressure, but also can be used to adjust the chamber pressure and the jack force immediately according to information feedback of ground settlement. Basic examples composed of combinations of different overburden depth and sandy ground are studied, and the validity of the model is discussed by comparing the results with simulation and Terzaghi's loosening earth pressure model.

**Keywords:** shield tunneling; earth pressure; ground settlement; soil-shield interaction.

## 1 Introduction

Shield tunneling has become a well-established tunnel construction method in various ground conditions. The shield has to overcome various complex loads during excavation, in which the earth pressure acting on cutter face is the main resistance [1-5]. Due to the disturbance of excavation, the earth pressure state in front of cutter face is destroyed, and the distribution of earth pressure is great singular and fluctuant. Thus the deviations of shield from the planned alignment and local collapse or blow-out of the ground near the tunnel face sometimes are occurred. To improve the construction accuracy and avoid failure of face stability, it is of major importance to study the earth pressure acting on cutter face.

During shield excavation, the chamber pressure and jack thrust are the most important operation parameters, especially for curve alignment excavation and automatic deviation correction operation. And the earth pressure acting on cutter face is considered to be the predominant factor affecting the shield operation, which direct

determines the chamber pressure. Moreover, the distribution of which determines the moment of deviation correction. Due to the large size of shield and complex working condition, such as geological diversity, multi-physics coupling, the earth pressure acting on cutter face cannot be given as a fixed value. Thus the studies of earth pressure focus on dynamic characters and the distribution.

Nowadays, the methods of studying earth pressure could be mainly classified into two categories: Rankine theory and limit equilibrium method. The first one is used to calculate the pressure with vertical earth pressure multiplied by lateral pressure coefficient. And vertical earth pressure is obtained by complete earth covering theory or loosening earth pressure theory. The former assumed that the vertical earth pressure is defined by the gravity of upper part soil of the shield. The latter is initially proposed by Terzaghi with considering the effect of shear strength [6-7]. The other one is applied to determine the earth pressure based on upper and lower bound theorem by a hypothetical sliding surface of soil in front of cutter face. Maidl B et al. proposed a two dimensional log spiral sliding surface to calculate the pressure acting on cutter face by solving moment equilibrium equation [8]. Davis et al. presented a three-dimensional local failure mechanism based on the upper bound theory to define the maximum height of tunnel heading which could be constructed under uniform support pressure in a given site [9]. Leca and Dormieux proposed a multi-block mechanism which is composed of some truncated rigid cones with circular cross sections and with same opening angles on the basis of Davis's work [10]. Multi-block failure mechanism suggested by Soubra is employed to investigate the possibility of progressive failure, especially the partial failure mechanisms, in extra large shield-driven tunnels [11]. Guilherm et al. presented a reliability-based approach for the three dimensional analysis and design of the face stability of a shallow circular tunnel driven by a pressurized shield [12]. Besides above mentioned, Mitsutaka Sugimoto proposed a model of load acting on cutter face with considering the nonuniformity of pressure distribution [4]. Sramoon A. et al. studied the distribution of earth pressure acting on cutter face by numerical simulation [5]. However, these models cannot represent the distribution of earth pressure acting on cutter face and the dynamic characters.

To solve above problems, this paper describes in detail the newly developed earth pressure model by establishing the relationship between the ground settlement and earth pressure. An object optimization function is proposed based on the principle of minimum potential energy. Furthermore, the validity of the model was discussed by comparing the results with test and models proposed by previous researcher.

## 2 Model of Earth Pressure

The stress of earth in front of cutter face is very complex three-dimensional nonlinear state. The orientation and magnitude of principal stress is varying with excavation. The global coordinate system is selected so that the  $x$  axis is vertically downwards; and the  $z$  axis is advancing direction.

Discrete elements are applied to simulate the soil in front of cutter face, and the loads acting on each element are composed of six forces: self-weight  $w_i$ , force on the upper

surface  $N_{i-1}$ , force on the lower surface  $N_i$ , and normal force  $N_{yi}, N_{zi}$ , the shearing force  $f_i$  on the element profile, as shown in Fig.1.

$$\begin{cases} N_{yi} = B\sigma_{yi}, N_{zi} = B\sigma_{zi} \\ N_{xi} = A\sigma_{xi}, f_i = 4\tau_i B \end{cases} \quad (1)$$

where  $B$  =the area of element profile,  $A$  = the area of element upper/lower surface,  $\tau$  =the shearing stress on the element profile,  $\sigma_y, \sigma_z$  =the stress in  $y, z$  direction,  $i$  =the number of elements.

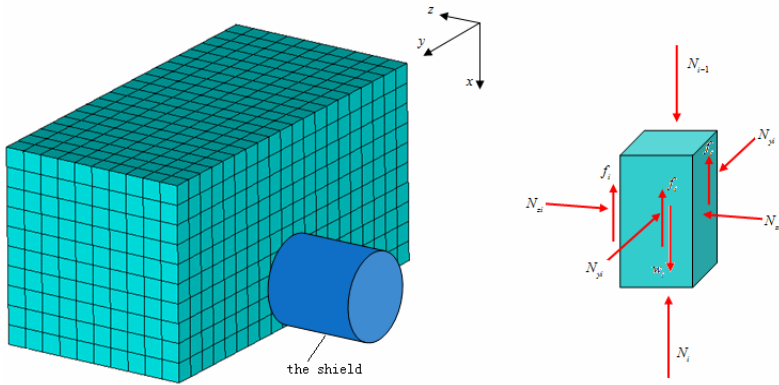


Fig. 1. The force diagram of soil element

Due to the ground settlement, the stress state of element could be considered as unloading process. Based on the Duncan-Chang nonlinear elasticity model, elasticity modulus  $E$  and volume modulus  $K$  can be represented by

$$E = K_{ur} p_a \left( \frac{\sigma_3}{p_a} \right)^n \quad (2)$$

where  $n$  =the test constant,  $p_a$  =the atmospheric pressure  $K_{ur}$  = the test constant.

To quantitatively study the earth pressure, some assumptions are necessary as follows:

1. The orientation of first principal stress  $\sigma_1$  is in accordance with the initial stress state during excavation. The displacement in  $x$  direction assumed to be  $u_x$ , according to generalized Hooker law,  $\sigma_{xi}$  can be calculated as follows:

$$\sigma_{li} = \sigma_{xi} = \gamma g h_i - E \frac{u_{xi}}{l} \quad (3)$$

where  $\sigma_{x0}$  = the initial vertical stress,  $l$  =the height of element,  $\gamma$  =the density of soil,  $g$  =the acceleration of gravity,  $h$  =the overburden depth.

2. The strains in  $y, z$  direction are neglected, the third principal stress  $\sigma_3$  is considered as:

$$\sigma_{3i} = K_o \sigma_{1i} \tag{4}$$

where  $K_o$  = the coefficient of earth pressure.

3. The shearing stresses  $\tau$  are equal on four profiles of each element. According to the equation of static equilibrium in  $x$  direction, the  $\tau$  can be obtained by

$$\tau_i = \frac{1}{4}(w_i + N_{i-1} - N_i) / B \tag{5}$$

4. The stresses in horizontal direction are equal. Based on Mohr's stress circle, the earth pressure acting on cutter face can be represented by

$$\begin{cases} \sigma_{yi} = \sigma_{zi} = \frac{\sigma_{1i} + \sigma_{3i}}{2} + \frac{\sigma_{1i} - \sigma_{3i}}{2} \cos 2\alpha_i \\ \sin 2\alpha_i = \frac{2\tau_i}{\sigma_{1i} - \sigma_{3i}} \end{cases} \tag{6}$$

where  $\alpha$  = the angle of normal direction of element profile to the first principal stress orientation.

Based on the above assumptions, a row element in vertical direction from the ground to the lower part of shield is studied; the earth pressure acting on cutter face is the function of  $u_{xi}$ .

### 3 Solution of Displacement Based on Nonlinear Optimization

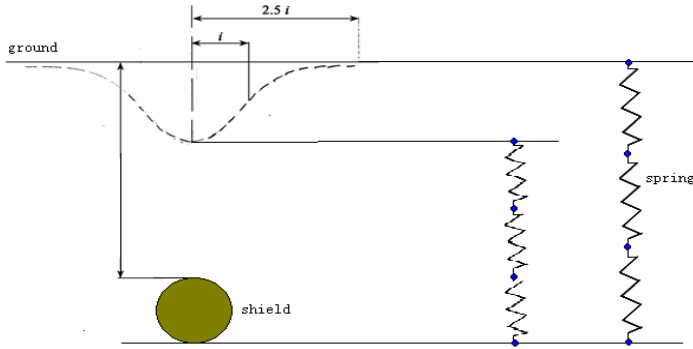
According to above analysis, the question is focus on finding a solution to the element displacement in  $x$  direction. The process of solving those parameters provides a reverse analysis of the ground settlement and the earth pressure. To predict the settlement induced by the tunneling, a formulation proposed by Peck can be represented as follows [13]:

$$\delta(y) = \frac{V_s}{\sqrt{2\pi}j} \exp\left(-\frac{y^2}{2j^2}\right) \tag{7}$$

where  $V_s$  = the ground formation loss,  $j$  = the coefficient of settling vessel breadth.

Thus the relationship between the ground settlement and the element displacement, as shown in Fig.2, can be represented by

$$\delta(y) = \sum_{i=1}^n u_{xi} \tag{8}$$



**Fig. 2.** Normal ground displacement around shield periphery (mm)

The ground settlement induced by cutter head is composed of the displacement of each spring in vertical direction during excavation. The principle of minimum potential energy is applied to find the solution of each element displacement. The gravitational potential energy and the deformation energy can be written as

$$U = w\delta / 2 - \frac{1}{2} \sum f_i x_i \quad (9)$$

$$V = \frac{A}{2} \sum (\sigma_{x_{0i}} + \sigma_{x_i}) u_{x_i} \quad (10)$$

where  $x_i$  is the sliding distance, can be calculated by

$$x_i = \sum_{j=i}^n u_{x_j} \quad (11)$$

Consequently, the optimization function can be defined as:

$$\begin{aligned} \min : \Pi &= \sum \Pi_i = V - U \\ \text{subject to} : \sum u_{x_i} &= \delta(y) \\ \sigma_{x_i} &\geq 0 \end{aligned} \quad (12)$$

Actually, this is a classical constrained nonlinear optimization problem. To solve the problem, the best approach is sequential unconstrained minimization technique (SUMT), which transforms the problem to unconstrained optimization problem by constructing a penalty function. The initial value for iterative algorithm can be given as inverse proportional to the overburden depth of each element. And the process of solving the displacement is carried out in the MATLAB.

## 4 Examples

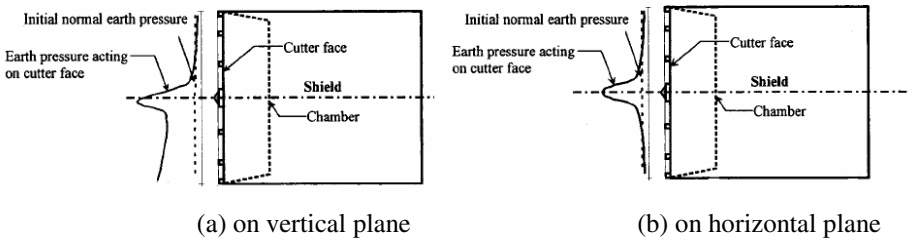
An earth pressure balanced shield with 6m diameter is studied in this paper. The basic examples are composed of combinations of different overburden depth, sandy ground.

**Table 1.** The calculation data[6]

test constant	the atmospheric pressure $p_a$	101325pa
	$n$	0.5
	$K_{ur}$	Calculated by 20m overburden depth and linear elasticity modulus
Sandy ground	density	1800kg/m <sup>3</sup>
	elasticity modulus $E$	25Mpa
	overburden depth	6m,9m,12m,18m
	The biggest ground settlement	176.8,128.9,89.57,20.99mm
	The coefficient of earth pressure	0.5
Structure	the diameter of shield	6m
	the area of element profile $B$	0.25
	the cross area of element $A$	0.25
	the height of element $l$	0.5

The results of numerical simulation, Terzaghi’s loosening earth pressure model and the proposed model proposed by this paper are compared. The calculation data are illustrated in Table 1. To ensure the comparability, the different parameters of these models are uniform as possible as far. Thus, the elasticity modulus and volume modulus used in the simulation are same as the calculation results of Duncan-Chang model in the 20m overburden depth, and the maximum ground settlement is also same as the simulation results.

Sramoon A. et al. studied the distribution of earth pressure acting on cutter face by numerical simulation, as shown in Fig.3 [5]. The bulged earth pressure at the center is caused by the penetration. In the horizontal plane, the earth pressure is nearly equality. On the other hand, the earth pressure in the vertical plane is nearly proportional to the overburden depth. Fig.4 illustrates the results calculated from the new model: the abscissa of Fig.4 (a) denotes the vertical direction from the shield top to the bottem and that of Fig.4 (b) denotes the horizontal direction from the shield left to the right side. And the same variation tendency is shown in the both directions compared with Fig.1 except the bulged at the center, which is not considered in new model.



**Fig. 3.** Normal earth pressure acting on cutter face

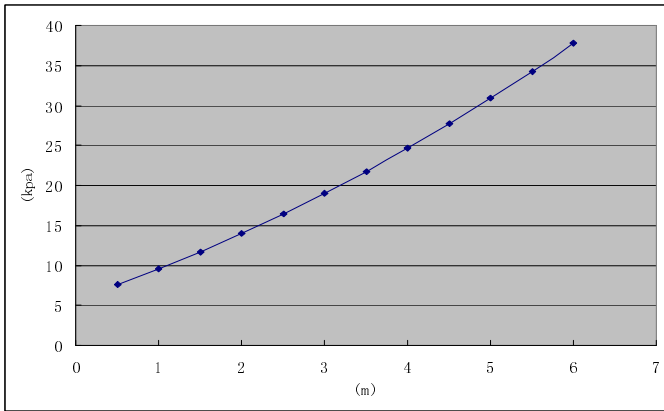


Table 2 shows the pressure ratio (the current earth pressure / the initial vertical earth pressure) and average pressure acting on the cutter face calculated by numerical simulation, Terzaghi's loosening earth pressure model and the new model, respectively. The following are found:

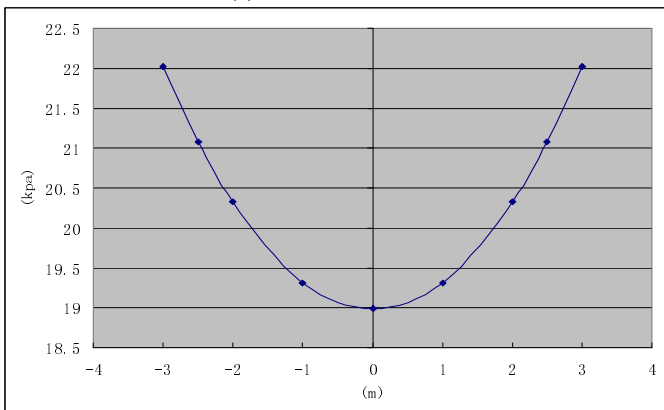
1. The values of latter two models are far more than that of simulation results, and that can be explained as the linear earth constitutive model for simulation.

2. The pressure ratio of new model increased with the overburden depth is coinciding with variation tendency of the soil shear strength. Whereas, the value of Terzaghi's model shows tendency in reverse. According to Xiaowen Zhou's research and centrifuge test [15], Terzaghi's model cannot be used for big overburden depth and there is more error compared with test.

3. For the 18m overburden depth, the pressure ratio of new model is close to initial coefficients of earth pressure. Therefore, the turbulence of soil induced by shield tunneling is less with the increasing of overburden depth in sandy ground.



(a) In vertical direction



(b) In horizontal direction

**Fig. 4.** The distribution of earth pressure acting on cutter face

**Table 2.** The results of earth pressure

Overburden depth (m)	Initial vertical earth pressure (kpa)	Simulation results		Loosening earth pressure model		New model	
		Pressure (kpa)	ratio	Pressure (kpa)	ratio	Pressure (kpa)	ratio
6	105.84	12.7	0.12	45.45	0.43	22.45	0.21
9	157.76	13.1	0.08	63.37	0.40	55.74	0.35
12	211.68	13.7	0.06	78.71	0.37	77.99	0.37
18	317.52	15.6	0.05	103.03	0.32	155.58	0.49

Through the above analysis, similar distribution of pressure acting on cutter face obtained from the new model is proved. The average earth pressure is valid and more reasonable from the engineering viewpoint.

## 5 Conclusions

The process of shield tunneling is complex interaction between the shield and the soil. In this paper, Duncan-Chang nonlinear elasticity model is applied to simulate the soil-shield interaction, and Peck's formula is used to establish the relationship between the ground settlement and earth pressure. Furthermore, an earth pressure model acting on cutter face is proposed for automation control systems based on the principle of minimum potential energy. Through the solving constrained nonlinear optimization problem, the results of examples are compared with simulation and Terzaghi's loosening earth pressure model. The results confirm that proposed model has similar precision and more reasonable in lower overburden depth. To great extent, the new proposed model not only can provide the distribution of earth pressure, but also can be used to adjust the chamber pressure and the jack force immediately according to information feedback of ground settlement. However, it is should be mentioned, there are some limitation, such as not considering the arching of soil, neglecting the horizontal ground displacement, and the validity should be examined in the shield tunneling test site. In the future, the model needs to be further investigated by simulation and experiments. Besides, three-dimensional model and dynamic factor are also worth researching.

## Acknowledgement

The authors acknowledge supports from the National Natural Science Foundation of China (50705035 and 50625516), the National Fundamental Research Program of China (973) (2007CB714000).

## References

1. Loganathan, N., Poulos, H.G.: Centrifuge model testing of tunnelling induced ground and pile deformations. *Geotechnique* 50(3), 283–294 (2000)
2. Lee, G.T.K.: The Effects of advancing open face tunneling on an existing loaded pile. *Journal of Geotechnical and Geoenvironmental Engineering* 131(2), 193–201 (2005)

3. Ueno, T.: Direction control system of shield tunneling machines. *Ishikawajima-Harima Engineering Review* 37(3), 203–218 (1997)
4. Sugimoto, M., Sramoon, A.: Theoretical model of shield behavior during excavation I: Theory. *Journal of Geotechnical and Geoenvironmental Engineering* 128(2), 138–155 (2002)
5. Sramoon, A., Sugimoto, M., Kayukawa, K.: Theoretical model of shield behavior during excavation II: Application. *Journal of Geotechnical and Geoenvironmental Engineering* 128(2), 156–165 (2002)
6. Konda, T., Ogasawara, M., Sasaki, K., Tsuno, K., Fujii, Y.: Study on design of a pair of shield tunnels situated very closely by centrifugal model test. In: *Proc., Int. Cong. on Tunnel and Underground Works, Tunnels for People*, pp. 151–156 (1997)
7. Nomoto, T., Imamura, S., Hagiwara, T., Kusakabe, O., Fujii, Y.: Shield tunnel construction in centrifuge. *Journal of Geotechnical and Geoenvironmental Engineering* 125(4), 289–300 (1999)
8. Maidl, B., Herrenknecht, M., Anheuser, L.: *Mechanized Shield Tunneling*, Berlin (1996)
9. Davis, E.H., Gunn, M.J., Mair, R.J., Seneviratne, H.N.: The stability of shallow tunnels and underground openings in cohesive material. *Géotechnique* 30(4), 397–416 (1980)
10. Leca, E., Dormieux, L.: Upper and lower bound solutions for the face stability of shallow circular tunnels in frictional material. *Geotechnique* 40(4), 581–606 (1990)
11. Soubra, A.-H.: Upper-bound solutions for bearing capacity of foundations. *Journal of Geotechnical and Geoenvironmental Engineering* 125(1), 59–68 (1999)
12. Mollon, G., Dias, D., Soubra, A.-H.: Probabilistic analysis and design of circular tunnels against face stability. *International Journal of Geomechanics* 9(6), 237–249 (2009)
13. Peck, R.B.: Deep excavations and tunneling in soft ground. In: *Proceedings of the 7th International Conference on Soil Mechanical and Foundation Engineering*, Mexico City, pp. 225–290 (1969)
14. Qin, J.: Study on face deformation and collapse of earth pressure shield tunnel. Hehai University, Nanjing City (2005)
15. Zhou, X.: Study on stability mechanism and relaxed soil pressure in sandy soil during excavation. *Journal of Yangtze River Scientific Research Institute* 16(4), 9–14 (1999)

# Incident Angle Measurement of Collimated Laser Based on the Pinhole Principle

Minghua Pan and Guoli Zhu

State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science & Technology, Wuhan, 430074, China,  
pmh24@163.com

**Abstract.** The incident angle of collimated laser is difficult to measure accurately within a large measurement range. A measurement method based on the pinhole principle, which is more accurate after modifying the lens distortion and focal length, was presented in this paper. The parameters which influence the accuracy, such as lens distortion coefficients and focal length, were analyzed using Monte Carlo method and corrected by experimental data to enhance the accuracy of angle measurement. The proto type based on the method was developed. Experiments show that the proto type gets an accuracy less than  $0.006^\circ$  in the entire measurement range of  $\pm 10^\circ$ , which is in accord with theoretical predictions. The reliability of the measurement accuracy was also proved via repetitive experiments. The proto type can be widely used in many fields such as automatic controls of industrial robot and large engineering machinery.

**Keywords:** incident angle; pinhole; lens distortion.

## 1 Introduction

Incident angle measurement of collimated laser has wide applications in different fields such as aero space[1,2], precision measurement[3], automatic controls[4] and so on. Over the last several decades various methods have been used to improve the optical systems of incident angle measurement. Those measurement methods represented by grating sensors, laser interference and autocollimator are mainly applied in precision measurement for their high-precision and small measurement ranges. For the complex structures of them are susceptible to interference, they can't be applied to poor working conditions or great changes of angles such as automatic construction surveying in large engineering machinery and industrial robot.

Compared with the methods mentioned above, the measurement model shown in figure 1 is easy to realize the measurement of the incident angle. The laser device automatically aims at the center of the corner cube reflector (CCR) and emits laser to measurement the center point. Most of the laser beam reflect to it but a little of them pass through the flat-top and then the lens focuses on the CCD area in the camera. Based on the pinhole principle, the incident angle can be acquired by

$$\alpha = F(u, v, f) = \tan^{-1}(\sqrt{u^2 + v^2} / f) \quad (1)$$

where  $f$  is the focal length and  $(u, v)$  is the coordinate of the spot center in the CCD area.

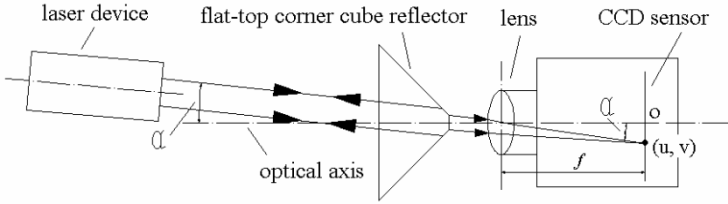


Fig. 1. The incident angle measurement model

## 2 Error Analysis and Correction of the Model

It is easy to know from equation 1 the measurement errors are caused by three main factors: the error of the focal length, spot deviation due to lens distortion and the error of spot center determination. Focal length error and lens distortion are systemic errors but the spot center determination error is accidental error. In recent years, many open literatures have discussed the technique of spot center determination. As known from them, location precision of laser spot reaches sub-pixel after image processing by special algorithm. Taking no consideration of the impact of lens distortion and focal length error, the theoretical precision of the incident angle can reach  $0.006^\circ$  when the nominal size of the focal length is 35mm and the dimension of the pixel is 7.4 $\mu$ m.

### 2.1 Lens Distortion Error Model

Research on the lens distortion has been going on for many years. Conrady introduced the camera lens distortion firstly in 1919. Then Brown[5,6] proposed the classical model of lens distortion containing radial distortion and tangential distortion in 1966. After that, dozens of modification methods based on the model have been reported[7-10] in recent years. Most of them depended on the high-precision chessboard and complex interactive algorithm.

For the purpose of analyzing the spot error resulted from the lens distortion, distortion model should be determined. In general, radial distortion constitutes the main part of the lens distortion, but tangential distortion can't be ignored in some situations. In the Brown model, lens distortion can usually be expressed as follows.

$$\begin{aligned}\Delta u &= u_d - u = (k_1 r^2 + k_2 r^4) + (2 p_1 u v + p_2 (r^2 + 2 u^2)) \\ \Delta v &= v_d - v = (k_1 r^2 + k_2 r^4) + (p_1 (r^2 + 2 v^2) + 2 p_2 u v)\end{aligned}\quad (2)$$

where  $u$  and  $v$  are the theoretical coordinates of image;  $r$  implies the distance from  $(u, v)$  to the center of the image plane;  $k_1$ ,  $k_2$ ,  $p_1$  and  $p_2$  are the coefficients of radial and tangential distortions. This model can explain most of the distortion and the next discussions are based on it.

### 2.2 Error Correction

As shown in formula 1 and 2, the systemic error of incident angle  $\alpha$  can be expressed as

$$\Delta\alpha = F(u_d(k_1, k_2, p_1, p_2), v_d(k_1, k_2, p_1, p_2), f + \Delta f) - F(u, v, f) \tag{3}$$

It is obvious that  $\Delta f$  leads to linear error and lens distortion results in nonlinear error. If the lens distortion is ignored,  $\Delta\alpha$  will be linearly amplified with the value of  $\alpha$ .

So the task of distortion modification is to eliminate the nonlinear error of  $\alpha$ . Formula 4 is used to characterize the linear degree of  $\Delta\alpha$ . There are  $n$  measurement points in the variation range and the errors of these measurement values are processed by linear regression. Objective function  $G$  is the quadratic sum of linear regression. The closer to the true values the distortion parameters are, the smaller the value of the function  $G$  will be. When  $\min G$  is acquired by nonlinear optimization, the values of distortion coefficients can be determined.

$$G(k_1, k_2, p_1, p_2) = \sum_{i=1}^n (\Delta\alpha_i - L(\Delta\alpha_i))^2 \tag{4}$$

Many literatures have discussed the nonlinear optimization algorithm to acquire the minimum of objective function. Most of them used Newton iteration or quasi-Newton iteration method to deal with the objective function. When the unknown parameters of objective function are more than one, gradient vector, Jacobin matrix and Hessian matrix should be calculated and applied to the iterative computation. If the expression of the objective function is very complex, those vectors and matrixes will be very difficult to deduce. More importantly, it needs to modify the matrix in each iteration for the inverse matrix of Hessian matrix may not exist in some iterative points.

As known from the above, Newton iteration and quasi-Newton iteration are inefficient and inapplicable when the objective function is complex and the second order partial derivatives are difficult to deduce. In order to avoid these defaults, difference iteration method is used in this paper which doesn't depend on the partial derivatives and is easy to be derived.

When difference iteration is adopted, the initial value of unknown parameters and the gradient vector of each iteration should be determined. The iteration expression is

$$k^{(i+1)} = k^{(i)} + t_i B^i p^i \tag{5}$$

In this formula,  $k$  is the vector of distortion coefficients;  $t_i$  is the step size;  $B^i$  is the matrix of iterative direction;  $p^i$  is the gradient vector in the iteration point.

The members of gradient vector in formula 5 are proportional to the sensitivity of the distortion coefficients in formula 3. In order to analyze the sensitivities of the distortion coefficients to the incident angle, the variances of the incident angle error are estimated by Monte Carlo simulations. The result of the analysis is shown in Fig 2 which indicates the significant linear relationships between the incident angle error and distortion coefficient errors. Therefore, when the nominal focal length of the lens is 35mm, the pixel size is 7.4um and the area of CCD is 1600×1200, the gradient vector  $p$  in formula 5 can be constantly set to

$$p = \begin{bmatrix} \nabla_{k_1} F(k_1, k_2, p_1, p_2) \\ \nabla_{k_2} F(k_1, k_2, p_1, p_2) \\ \nabla_{p_1} F(k_1, k_2, p_1, p_2) \\ \nabla_{p_2} F(k_1, k_2, p_1, p_2) \end{bmatrix} = \begin{bmatrix} 0.0388 \\ 1 \\ 0.00416 \\ 0.00302 \end{bmatrix} \tag{6}$$

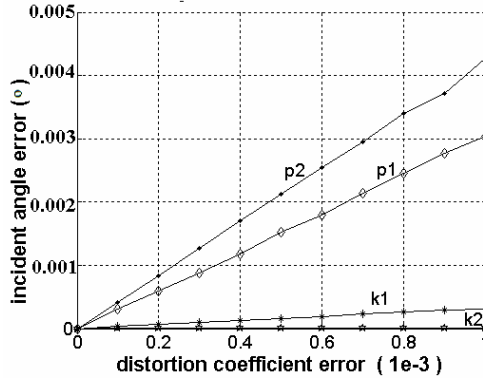


Fig. 2. The sensibility curves of distortion coefficients

The step size  $t_i$  and matrix  $B^i$  express the optimal step and iterative direction of the distortion coefficients vector. Along the iterative direction, the value of objective function  $G$  decreases fastest in the current iteration. For the gradient vector is constant, the matrix  $B^i$  has 16 optional directions described as

$$B = \begin{bmatrix} (-1)^{m_1} & & & \\ & (-1)^{m_2} & & \\ & & (-1)^{m_3} & \\ & & & (-1)^{m_4} \end{bmatrix}; m_1, m_2, m_3, m_4 = 0, 1 \tag{7}$$

As stated above, the iteration process can be shown in Fig.3. The point sets of  $H$  are collected by measuring the incident angle. Points sets of incident angle error can be acquired by the theoretical points sets  $\Gamma$  and the measuring points sets  $\gamma$ . The optimization objective is to acquire the minimum value of  $G$  that is the quadratic sum of linear regression. The initial value of the distortion coefficients vector  $k$  is set to  $[0 \ 0 \ 0]^T$ . The iteration process is carried out as follows:

1.  $k(i) \Rightarrow k, G(i) \Rightarrow G$ ;
2. determine the optimal step size  $t^*$  and  $B$  to minimize the objective function  $G$ ;
3. if  $G(i+1) < G(i)$ , to the next step, otherwise revert the value of  $k$  and  $G$ , back to step 2;
4. if the absolute difference value between  $G(i)$  and  $G(i+1)$  is less than the allowable error  $\xi$ , finish the iteration, otherwise back to step 1.

Using the distortion coefficients vector  $k$  acquired from the iteration, nonlinear error of the incident angle  $\alpha$  is eliminated, but the linear error owing to  $\Delta f$  still exists. Least squares approach can be applied to acquire the correct value of focal length  $f$  to minimize the value shown in equation 8. Then, the final error of the incident angle is the accidental error caused by the spot center determination.

$$R(f) = \sum_{i=1}^n (\Delta \alpha_i)^2 \tag{8}$$

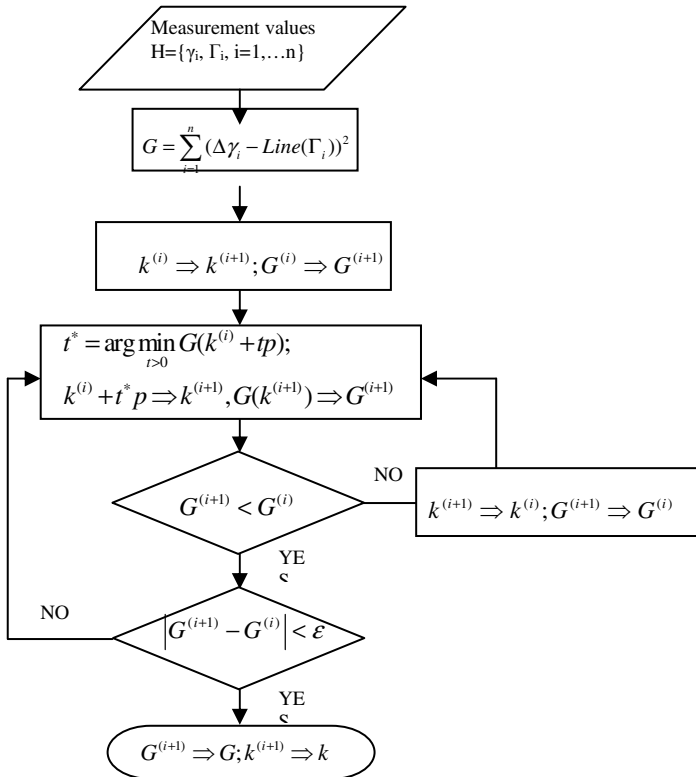


Fig. 3. The flow of iterative algorithm

### 3 Experimental Device and Data Processing

Fig.4 shows the experimental device of the incident angle measurement. CCR is installed in front of the camera. Both of them are placed on a rotary platform which is driven by a high precision servo motor with an absolute encoder in the rear. The absolute accuracy of the servo motor reaches  $0.005^\circ$  and can be applied to define the variation value of the angle. Laser device is placed far away from them and locks the center of CCR. When the turntable turns to a certain direction, the focal spot will move along a line. The coordinates of the trajectory of the focal spot in the imagine plane are collected to calculate the variation of the incident angle. At the same time, the angle of the servo motor and the azimuth angle of the collimated laser are read to calibrate the theoretical value of the incident angle. After a series of measuring points are collected and calculated by formula 1 and 2, set H is acquired. The next step is to deal with these data by the iterative algorithm.

For the coefficients vector describes the image distortion in the two dimensions of the whole plane, the focal spots should be distributed as broad as possible. The two lines illustrated in Fig.5 are the trajectory of focal spot which cover the full scale of



the two dimensions. The zero positions of the incident angle in the two trajectory lines are  $O'$  and  $O''$  respectively, which are the projective points from the principal point to the trajectory lines.

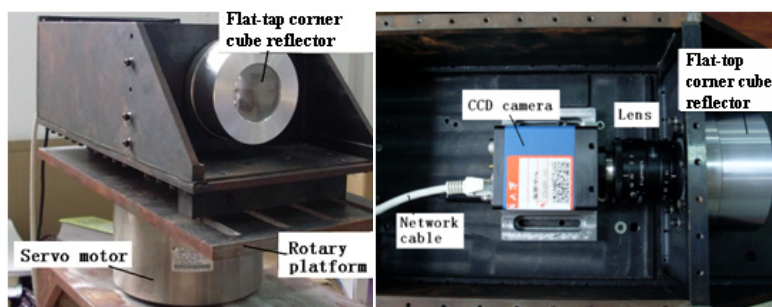


Fig. 4. Experimental devices

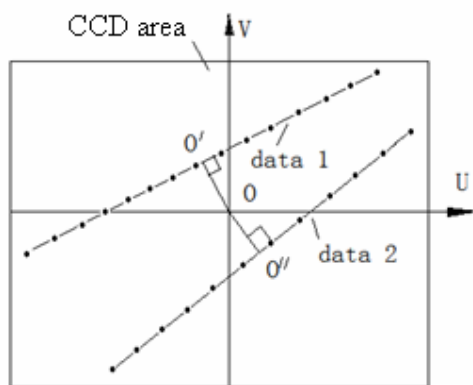


Fig. 5. Spots trajectory in the CCD area

Incident angle errors calculated by these coordinates are shown in curve I of Fig.6. It's easy to draw a conclusion that the distortion of camera lens leads to obvious nonlinearity. After data processing by the iterative algorithm, the optimal solutions of the distortion coefficients are as follows:  $k_1 = -0.18601$ ,  $k_2 = -0.074$ ,  $p_1 = -0.00167$ ,  $p_2 = 0.000906$ , which are used to modify the lens distortion. The effect of modification is shown in curve II. After that, the correct value of  $f$  which is acquired by least-squares iterative as 35.316mm. Curve III shows the last modified results which distribute randomly near the zero point.

In order to verify the effect of the modification, other experiments are carried out. The results are shown on Fig.7. It is obvious that most of the measurement errors are little than  $0.006^\circ$ . The results are valid and reliable in the whole variation range of the incident angle and accord with the theoretical precision.

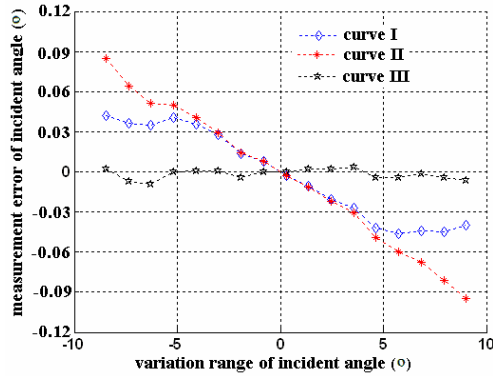


Fig. 6. Incident angle modification curves: (I) initial data; (II) data after distortion modification (III) data after distortion modification and focal length correction

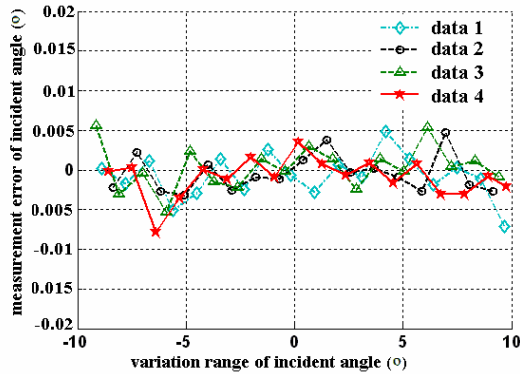


Fig. 7. Repetitive experimental results curves

### 4 Conclusion

A method for the incident angle measurement of the laser is presented in this paper. This method is based on the pinhole principle and Brown distortion model. After analysis by Monte Carlo method, the sensitivities of different distortion coefficients are deduced. The calibration and modification process of the lens distortion are greatly simplified by uniting the four distortion coefficients into one vector. The experiment proves the superior effect of the correction method. The repeatability experiments shows that absolute errors of the incident angle are little than 0.006° which are in accord with the theoretical predictions.

**Acknowledgments.** This work is supported by the National Basic Research Program of China (No. 2007CB714005).

## References

1. Buonocore, M., Grassi, M., Rufino, G.: APS-based miniature sun sensor for earth observation nanosatellites. In: *Acta Astronautica*, 4th IAA International Symposium on Small Satellites for Earth, vol. 56(1-2), pp. 139–145 (2005)
2. Kouzmin, V. S., Cherem Oukhin, G.S., Fedoseev, V.I.: Miniature sun sensor. In: *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 2739, pp. 407–410 (1996)
3. Qin, S., Chen, Y., Su, Y., et al.: High precision automatic angle measuring system. *Instrument Technique and Sensor* (11), 13–15 (1999) (in Chinese)
4. Pan, M., Ren, Y., Mao, S.: The real-time detection of pose and control of laser guiding vehicle (LGV). *Journal of National University of Defense Technology* 16(2), 34–37 (1994) (in Chinese)
5. Devernay, F., Faugeras, O.: Straight lines have to be straight. *Machine Vision and Applications* 13, 14–24 (2001)
6. Devernay, F., Faugeras, O.D.: Automatic calibration and removal of distortion from scenes of structured environments. In: *SPIE*, San Diego, CA, vol. 2567 (1995)
7. Tsai, R.: A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-self TV camera lenses. *IEEE Journal of Robotics and Automation RA-3*, 323–344 (1997)
8. Weng, J., Cohen, P., Herniou, M.: Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14, 965–980 (1992)
9. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 1330–1334 (2000)
10. Devernay, F., Faugeras, O.: Straight lines have to be straight. *Machine Vision Applications* 13, 14–24 (2001)

# A Novel Design for the Self-reconfigurable Robot Module and Connection Mechanism

Enguang Guan, Weixin Yan, Dongsheng Jiang, Zhuang Fu, and Yanzheng Zhao

School of Mechanical Engineering, Shanghai Jiao Tong University, State Key Laboratory of Mechanical System and Vibration, Shanghai, P.R. China

{EnguangGuan, WeixinYan, DongshengJiang, ZhFu, Yzh-Zhao}@Sjtu.edu.cn

**Abstract.** Many researchers have paid more and more attention to the lattice self-reconfigurable modular robot for its excellent flexibility in connection and separation movements of modules. The paper proposes a novel design of self-reconfigurable robot module (M-Lattice) with the overall description. Moreover, the key features of module's genderless pin-slot-based connection mechanism are introduced. Finally, the connection mechanism is proved to be feasible for the tasks of module's self-reconfiguration by the experiments.

**Keywords:** Self-reconfigurable robot, Modular, Connection mechanism.

## 1 Introduction

As a front area of research, self-reconfigurable modular robot has got more and more attention all over the world recently with the advantages of the reconfigurable robot and the mobile robot, the modular robot is a typical autonomous mechanical unit which has abilities of independent movement, computation and communication[1-2]. A large number of modules with the same functions constitute a self-reconfigurable modular robot system. The most prominent feature of the system is the ability to change its shape without outside help through the module separation between the active connections or to change the system topology in order to adapt to the complex non-structural work environment. This target-based self-reconfigurable ability is given to self-reconfigurable modular robot with the flexibility that a fixed structure robot can not be achieved. That makes it effective in an unknown environment, such as disaster search operations, space or ocean explorations and other areas [3-4].

For the design of self-reconfigurable robot module, connection mechanism is a particularly important part for a flexible and reliable connection mechanism has laid the foundation for the self-reconfigurable movements among modules [5]. Up to now, many researchers have built several prototypes. The adopted connection mechanism can generally fall into such types as mechanistic and electromagnetic. A typical case of mechanical interface is Superbot modules designed by Wei-Min Shen etc. from USC (University of Southern California), which use a multi-jaw structure to accomplish connection and separation of modules by imitating human hands [6]. M-TRAN II, designed by Yoshida etc of AIST (Advanced Industrial Science and Technology) applies an electromagnetic interface [7]. The magnet on its plane is used

for connection and a shape memory alloy (SMA) provides the driving force to separate when it gets heat.

The domestic research in this field is also making a continuous progress. HitMSR robot modules (Jie Zhao, et al, Robotics Institute of Harbin Institute of Technology) uses an electromagnetic interface[8], and the M-Cubes (Yanqiong Fei, et al, Research Institute of Robotics, Shanghai Jiaotong University) adopts a mechanical interface with a pin-hole-based design [9].

Compared to electromagnetic interface, mechanistic one is a more reasonable for the self-reconfigurable module owing to its remarkable advantages of being more efficient, reliable and energy-saving. On the other hand, it requires more parts with a higher machining accuracy. And too many moving parts will increase the cost and make it too complicated. In order to improve these problems, a novel mechanistic connection mechanism applied to a three-dimensional self-reconfigurable robot module M-Lattice is proposed in this paper.

## 2 Design of the M-Lattice Module

### 2.1 Overall Design

The topology of the M-Lattice module is lattice structure just as shown in Fig.1. Thus the self-reconfigurable system constituted by the modules looks like a mesh. Each node in the mesh represents a robot module and the changes of nodes` positions mean the system structure transformations. This kind of lattice structure brings great convenience to realize the self-reconfigurable movements because a module in the mesh can move from one node to the others neighboring [10-11]. As illustrated in Fig.1, the robot module designed to be a center-symmetric three-arm structure, consists of a triangular central frame and three identically constructed mechanical arms which are installed on the three side walls of the center frame. Each mechanical arm contains two joints and one connection mechanism. The connection, separation and other movements of modules are completed by three degrees of freedom arms. Therefore, the connection mechanism at the end of the arm is essential to the stability

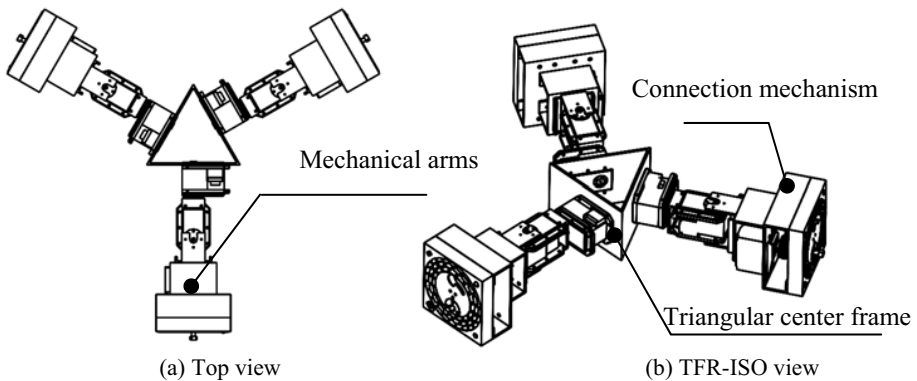
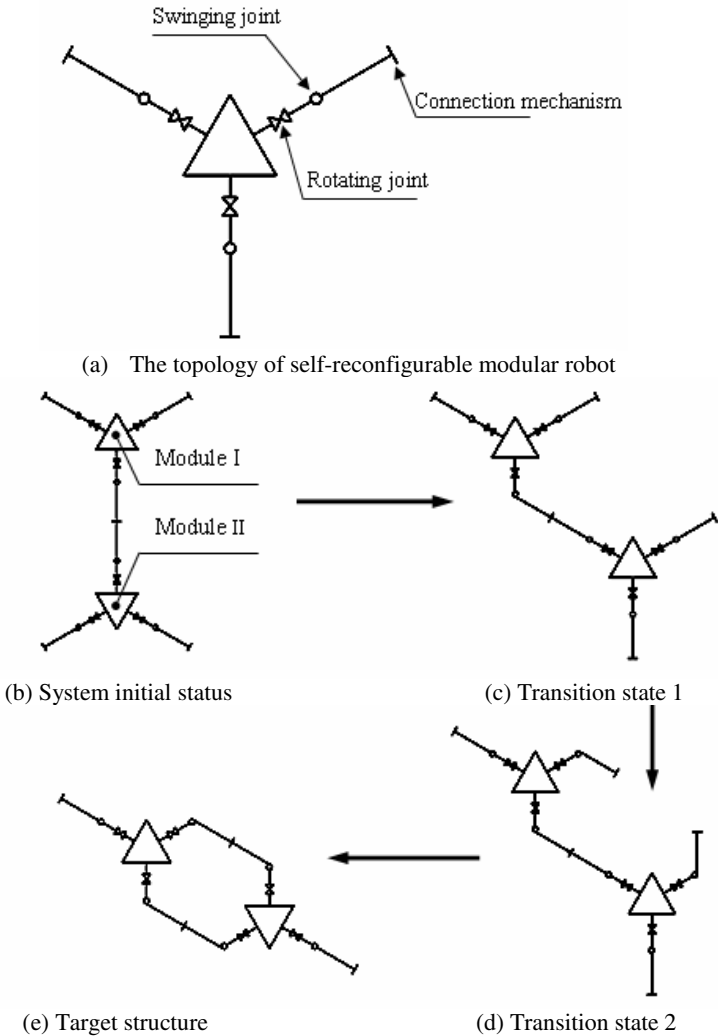


Fig. 1. A novel lattice self-reconfigurable robot module

and reliability of the self-reconfigurable module system. The connection mechanism proposed in this paper is a genderless pin-slot-based structure with a pin and a rocking slot which are designed on its end surface. The pin is inserted into the rocking slot and then rotated oppositely to make the two modules conjoint tightly when there is a need for connection. The connection mechanism possesses good symmetry of space, thus the modules get the same functions and can be replaced mutually.

### 2.2 The Configuration of the Mechanical Arms

The mechanical arms use servomotors as the actuators to drive the joints. There are two degrees of freedom (one rotating joint and one swinging joint) on the forepart of



**Fig. 2.** The self-reconfiguring process of two modules system

the single arm at least for the purpose of flexible positioning of the connection mechanism within its workspace. Consequently, at least three degrees of freedom must be configured in one mechanical arm after considering the rotation degree of freedom in the connection mechanism. Based on the topological structure of the self-reconfigurable modular robot, the self-reconfiguring process in two modules system is shown in Fig.2.

If some extra degrees of freedom are added like swinging joint into the beginning section of the arm as the fourth degree of freedom, the moving ability among modules will be greatly enhanced and they can even realize some rather complex movements like multi-legged walking by self-reconfiguration. But the more degrees of freedom of the module configures, the higher requirement is in the system driving and controlling. Therefore, the first generation prototype of the self-reconfigurable modular robot M-Lattice adopts the three degrees of freedom structure.

### 2.3 The Pin-Slot-Based Connection Mechanism

An accurate and reliable connection mechanism is the basis for the self-reconfigurable movements among modules. The connection mechanism referred in this paper is a novel pin-slot-based style as shown in Fig.3.

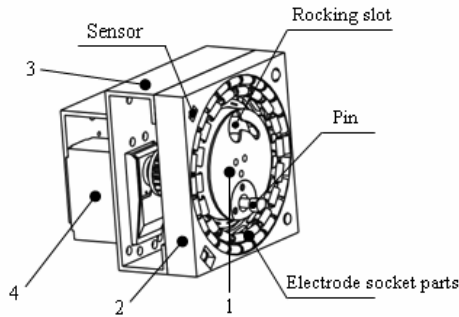


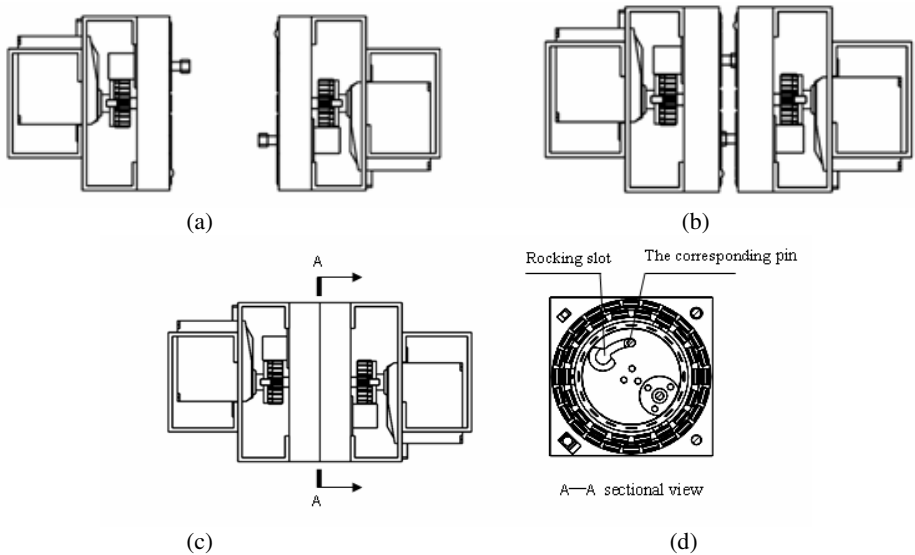
Fig. 3. The 3D virtual model of the connection mechanism

The connection mechanism includes four parts, i.e. the rotating block 1, the fixed base 2, the connecting frame 3 and the motor driving system 4. The rotating block embedded in the fixed base is the executor of the connection with a new feature that a pin and a rocking slot are both specially designed on the end surface of the mechanism. There is a Hall sensor, an infrared sensor and the electrode socket parts used to communicate among modules in a circular hole of the fixed base. The connecting frame is used for joining the fixed base to the mechanical arms. A driving motor fixed on the connecting frame provides the rotating driven force through a gear.

The two connection mechanisms keep far away from the other at the initial state when the connection movement is ready to proceed (Fig.4.a), then they get close

gradually in order to align through the arms' position controlling information from the Hall, infrared sensor and the gyro sensor in the module center. At the same time while the connection surfaces get alignment, the pin plugs into corresponding rocking slot with the help of the bugle shape feature at the slot entrance (Fig.4.b). And then the servomotors drive the rotating blocks of the connection mechanisms relatively. In the rotation, the pin slides along the rocking slot and finally joints tightly at the end of slot (Fig.4.c & Fig.4.d). The head feature of the pin and the ending shape of the rocking slot are both specially designed to make them coincident with the purpose of positioning and fastening between them. Moreover the axial movement is restricted by a cover plate to prevent the pin from skidding off the rocking slot axially. The homogeneous structures of the connection mechanisms make the pins insert into the corresponding rocking slots synchronously. The connection and the separation movements can be executed by the servomotor while rotating clockwise and anticlockwise. The connecting procedure is shown in Fig.4.

The driving motor's stopping torque is  $37.7\text{kgf}\cdot\text{cm}$  at 16V voltage and the gear reduction ratio is 1. The reasons of adopting this gear transmission are following: the assembly processability of structure is improved and there is no need to increase the gear reduction ratio to obtain more powerful torque output since the driving motor is capable for the system load requirement.



**Fig. 4.** The illustration of the process of the connection mechanisms

## 2.4 The System Error Analysis

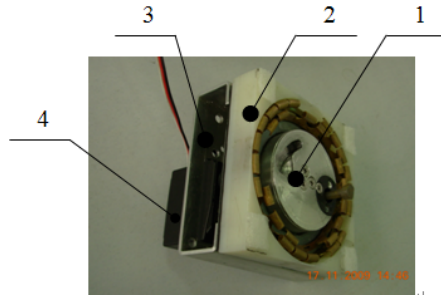
The errors which affect the system's motion mainly come from three aspects i.e. the torque error of driving motors, the position one of sensors, and the systematic one of mechanical structure itself. However, those errors can be compensated through



different ways in the design process. First of all, servo motors with high accuracy are chosen to join driving parts. Inside the motors, closed-loop control circuits for position and velocity are integrated, which can achieve the motion control by receiving the command of the speed and angle from the central controller through Rs-485 Bus. Also, the central controller can monitor the speed and angular displacement through Rs-485 Bus to achieve higher position accuracy. Secondly, when the two connecting interfaces are far away from each other, the gyros installed on the central frame of the modular robot provides the robot's position and posture control with information of navigation. When the two connecting interfaces are close to each other, the infrared and Hall sensors, which are located diagonally on each surface, can further provide information for connecting orientation and give feedback of whether the two interfaces are nearly aligned. Furthermore, a kind of bugle structure for compensation is introduced in the design, which can automatically calibrate the motion error produced in the connecting process and ensure the right connection of the pin-slot mechanisms.

### 3 Experiments

The photo of the connection mechanism for the space self-reconfigurable modular robot (The first generation of M-Lattice) is shown in Fig. 5.



**Fig. 5.** The photo of the connection mechanism

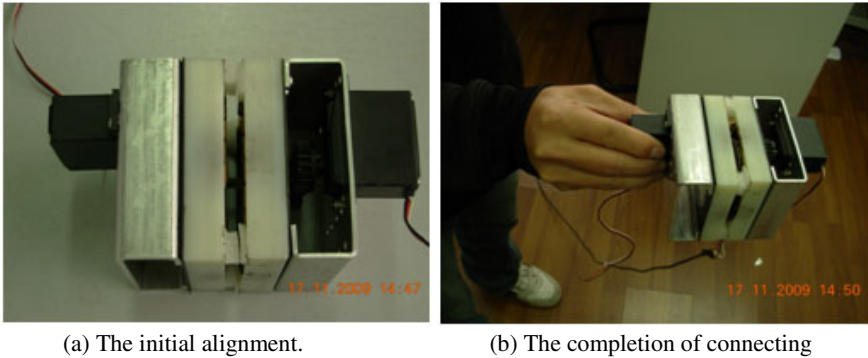
As illustrated in Fig.5, the connection mechanism includes four main parts, i.e. the rotating block 1, the fixed base 2, the connecting frame 3 and the motor driving system 4. For the connection mechanism two experiments are carried out.

#### 3.1 The Static Connecting Experiment

In the first step, the two separate connection mechanisms are placed on the experimental table with a good driving motors wiring. Then, the two connection mechanisms are manually aligned and the motors are driven through the motion controller (Fig. 6.a). So the two interfaces of the mechanisms are rotated and docked together by the pins. The entire procedure is executed smoothly and rapidly as the

time cost of the connecting movement is approximately 2 seconds. Afterwards the test shows that the strength of the connections between the two mechanisms is stable (Fig. 6.b). After connected, two modules will stay together, even though the power is cut off.

The diameter of Pin head is 7mm and the bugle entrance located in rocking slot's starting point gets 16mm in diameter. Therefore the maximum tolerate in offsetting distance between the pin's axis and the reference axis is 8mm—value of the bugle entrance's radius. Once the pin reaches the range of the bugle entrance (there could be some position error and in reality it always does), this bugle structure can provide sustained position correcting automatically for the pin and makes it finally get the expected location.



**Fig. 6.** The static connecting experiment

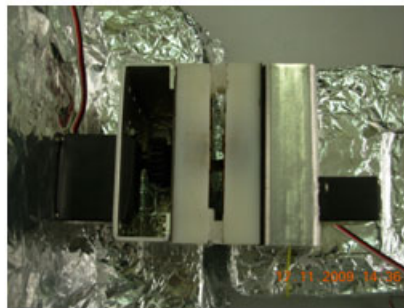
### 3.2 The Dynamic Connecting Experiment

A set of connection mechanism is placed on a static test platform (see Fig.7.a left), and the other unit is installed on a small toy car that can move along the horizontal direction (see Fig.7.a right). In the experiment, the relative position of the connecting interfaces can be regulated by controlling the motion of the car so that the two sets of the connection mechanisms are gradually close together, until the pins are inserted into the corresponding slots (Fig. 7.b). And then the motors are driven to complete connection. After the two devices are connected together, the small car takes the two mechanisms back to its initial place on the table (See Fig. 7.c). Finally, the motor is controlled along the other direction and makes the two devices disconnect. All the movements including alignment and connection take approximately 15 seconds which are mostly spend on controlling the toy car to align manually as the connection time cost is within 2 seconds. Based the well aligned situation by controlling the small car accurately, the connection can be accomplished every time. Through the experiment, the dynamic connecting method of the self-reconfigurable motion for the modular robots is illustrated.

The experiments show that the connection devices can achieve the module's connecting and separation movements successfully.



(a) The initial status



(b) The alignment status



(c) The car takes the connected devices

**Fig. 7.** Dynamic connecting experiment

## 4 Conclusions

This paper presents a novel space self-reconfigurable modular robot, which has a special connecting mechanism with pin-slot parts and hermaphrodite. In this paper, a self-reconfigurable module M-Lattice and the movement process of its connecting mechanism are described. The 3D virtual simulation and the experiments have proved that the connecting mechanism is simple and reliable. It can complete the connection and separation of the self-reconfigurable modular robots.

**Acknowledgments.** This work was partially supported by the National Natural Science Foundation of China under Grant No. 60875058 and SMC.

## References

1. Bing, W., Zhen, J.: Review on the status and development of modular reconfigurable robot technology. *Mechanical & Electrical Engineering Magazine* 25(5), 1–4 (2008)
2. Bin, L., Li-min, M., Zhi-hong, S., Ge, C.: Existing state of development for overseas modular robot with self-variable structure. *Journal of Machine Design* 22(5), 1–4 (2005)
3. Ping, X., Xin-jian, Z., Yan-qiong, F.: Design and Motion Planning Method of a Novel Modular Self-Reconfigurable Robot. *China Mechanical Engineering* 17(15), 1549–1552 (2006)
4. Liang, Z., Jie, Z., He-gao, C.: Research on hardware design and self-reconfiguration planning of self-reconfigurable robot. *Journal of Southeast University (Natural Science Edition)*, 35(2), 243–247 (2005)
5. Shen, W.-M., Kovac, R., Rubenstein, M.: SINGO: A single-end operative and genderless connector for self-reconfiguration, self-assembly and self-healing. In: *IEEE International Conference on Robotics and Automation*, pp. 4253–4258 (2009)
6. Salemi, B., Moll, M., Shen, W.-M.: SUPERBOT: A Deployable, Multi-functional, and Modular Self-Reconfigurable Robotic System. In: *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, October 9-15, pp. 3636–3641 (2006)
7. Murata, S., Yoshida, E., Kamimura, A., et al.: M-TRAN: Self-Reconfigurable Modular Robotic System. *IEEE/ASME Transactions on Mechatronics* 7(4), 431–441 (2002)
8. Yu-hua, Z., Jie, Z., Liang, Z., et al.: Novel Modular Self-reconfigurable Robot System. *Chinese Journal of Mechanical Engineering* 42, 175–178 (2006)
9. Yan-qiong, F., Qing-lei, D., Ping, X.: Design and Description of the Module in a Self-Reconfiguring and Self-Repairing Machine. *Journal of Shanghai Jiaotong University* 40(3), 448–450 (2006)
10. Murata, H., Kurokawa, S.: Self-Reconfigurable Robots-Cellular robots that transform their shapes. *IEEE Robotics and Automation Magazine*, 71–78 (2007)
11. Yim, M., Shen, W.-M., Salemi, B., et al.: Modular Self-Reconfigurable Robot Systems. *IEEE Robotics & Automation Magazine*, 1–11 (2007)

# The Forward Kinematics Analysis of 6-3 Stewart Parallel Mechanisms

Shili Cheng<sup>1</sup>, Hongtao Wu<sup>1,\*</sup>, Chaoqun Wang<sup>1,2</sup>, Yu Yao<sup>1,3</sup>, and Jianying Zhu<sup>1</sup>

<sup>1</sup> College of Mechanical and Electrical Engineering,  
Nanjing University of Aeronautics and Astronautics,  
Yudao Street.29, 210016 Nanjing, China

<sup>2</sup> College of Engineering,

Nanjing Agricultural University,  
Dianjiangtai Road. 40, 210031 Nanjing, China

<sup>3</sup> College of Aerospace Engineering,  
Nanjing University of Aeronautics and Astronautics,  
Yudao Street.29, 210016 Nanjing, China

mehtwu@126.com

**Abstract.** In order to study the forward kinematics analysis of 6-3 Stewart parallel mechanisms, this paper presents a new analytical method. Analyzing the coupling relationships between the position and orientation variables of the moving platform, additional useful information is obtained, thus eleven compatible algebra equations used to solve this problem are obtained. Based on this, the orthogonal complement method has been used several times to eliminate the redundant variables, the forward kinematics of 6-3 Stewart parallel mechanisms can be expressed as an eight order equation in the end. Finally a specific example is introduced to verify the method offered in this paper.

**Keywords:** Stewart Platform, Parallel Mechanisms, Forward Kinematics Analysis, Orthogonal Complement Method, Analytical Method.

## 1 Introduction

The 6-3 Stewart parallel mechanism is a kind of typical 6-6 Stewart parallel mechanism, the hinge connection points of the moving platform forms a triangle, and the hinge connection points of the static platform forms a hexagonal. The forward kinematics is to determine the position and orientation variables of the moving platform relative to the reference frame fixed in the static platform when the lengths of the links are given [1].

In the past decades, this problem has been paid more and more attention. Professor Liang presented an analytical method to solve this problem in [1]-[3]. He had successfully expressed it as a 16<sup>th</sup> order algebraic equation. Nanua [4] presented a 16<sup>th</sup> order algebraic equation in the result of the forward kinematics for the 6-3 Stewart parallel mechanism. This is not only a major breakthrough on this issue, but also the forward

---

\* Corresponding author.

kinematics analysis of the parallel mechanism. Akçali [5] presented a novel approach to solve this problem, but the method is too complexity. Song [6] presented a closed-form solution of this problem based on the Tetrahedron Approach. But the equation is not with respect to position and orientation variables. Many scholars attempt to solve this problem by the use of numerical methods too in [7]-[9], some progress had been achieved.

As the analytical method is more suitable for theoretical analysis, it is necessary to study it further more deeply. Referenced to [3], [10], this paper presents a new method to solve this problem. Through analyzing the coupling relationship between the position and orientation variables of the moving platform, 11 compatible equations which can be used to solve this problem are obtained. These additional information is very useful. Based on this, orthogonal complement method is used to eliminate the redundant variables, the forward kinematics of the 6-3 Stewart parallel mechanism can be expressed as an 8 degree algebraic equation.

This paper is organized as follows. Section 2 presents the basic concepts of the forward kinematics. Section 3 presents the 11 compatible equations. Section 4 orthogonal complement method is used to derive the 8 degree algebraic equation. In section 5, a numerical example is presented. Finally, in section 6, we present the conclusions.

## 2 Basic Concepts of the Forward Kinematics

Fig.1 shows a 6-3 Stewart parallel mechanism. For the convenience of analysis, the reference system  $OXYZ$  is attached to the static platform and its origin coincides with the joint  $B_1$ , the reference system  $O'X'Y'Z'$  is attached to the moving platform and its origin coincides with the joint  $A_1$ .

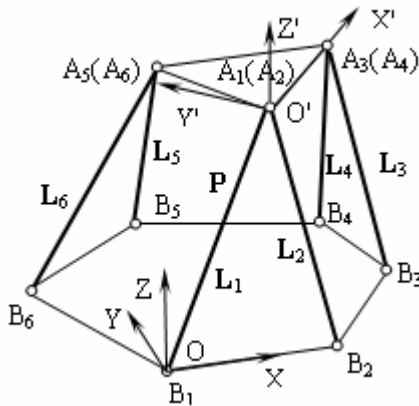


Fig. 1. a 6-3 Stewart parallel mechanism

The position of the joint  $B_i$  with respect to  $OXYZ$  can be expressed as:

$$B_i = [a_i \ b_i \ 0]^T \quad (i=1,2,\dots,6) \tag{1}$$

And the position of the joint  $A_i$  with respect to  $O'X'Y'Z'$  can be expressed as:

$$A_i = [p_i \ q_i \ 0]^T \quad (i=1,2,\dots,6) \tag{2}$$

The position vector  $\mathbf{P} = [x \ y \ z]^T$  describes the origin of  $O'X'Y'Z'$  with respect to  $OXYZ$ . The rotation matrix  $\mathbf{R}$ , which describes the orientation of the reference  $O'X'Y'Z'$  with respect to  $OXYZ$ , can be expressed as:

$$\mathbf{R} = \begin{bmatrix} l_x & m_x & n_x \\ l_y & m_y & n_y \\ l_z & m_z & n_z \end{bmatrix} \quad (3)$$

According to the orthogonality of the rotation matrix, there exist the following 3 equations:

$$l_x^2 + l_y^2 + l_z^2 = 1 \quad (4)$$

$$m_x^2 + m_y^2 + m_z^2 = 1 \quad (5)$$

$$l_x m_x + l_y m_y + l_z m_z = 0 \quad (6)$$

The vector of every link is:

$$\mathbf{L}_i = \mathbf{P} + \mathbf{R} \mathbf{A}_i - \mathbf{B}_i \quad (7)$$

From the above equations, we have:

$$\begin{aligned} & (p_i l_x + q_i m_x + x - a_i)^2 + (p_i l_y + q_i m_y + y - b_i)^2 \\ & + (p_i l_z + q_i m_z + z)^2 = l_i^2 \quad (i = 1 \sim 6) \end{aligned} \quad (8)$$

It can be noticed that there are only 9 unknowns in Eq.(8). The task is to solve the nine unknown variables. For the convenience of analysis, two intermediate variables  $w_1$ 、 $w_2$  are introduced in the following.

For  $i = 1$ , because of  $a_1 = b_1 = p_1 = q_1 = 0$ , Eq.(8) becomes:

$$x^2 + y^2 + z^2 = l_1^2 \quad (9)$$

For  $i = 2 \sim 6$ , eq.(8) can be simplified as:

$$p_i w_1 + q_i w_2 - a_i x - b_i y - C_i m_x = A_i l_x + B_i l_y + D_i m_y + E_i \quad (10)$$

where  $w_1$  and  $w_2$  are the two intermediate variables and defined as follows:

$$w_1 = l_x x + l_y y + l_z z \quad (11)$$

$$w_2 = m_x x + m_y y + m_z z \quad (12)$$

Their physical meaning will be presented later. And  $A_i$ 、 $B_i$ 、 $C_i$ 、 $D_i$ 、 $E_i$  are constants as follows:

$$\begin{aligned} A_i &= p_i a_i, B_i = p_i b_i \\ C_i &= q_i a_i, D_i = q_i b_i \\ E_i &= \frac{1}{2} (l_i^2 - l_1^2 - a_i^2 - b_i^2 - p_i^2 - q_i^2) \end{aligned} \quad (13)$$

From those above equations, we can get the following five equations:

$$w_1 = c_{1,1}l_x + c_{1,2}l_y + c_{1,3}m_y + c_{1,0} \quad (14)$$

$$w_2 = c_{2,1}l_x + c_{2,2}l_y + c_{2,3}m_y + c_{2,0} \quad (15)$$

$$x = c_{3,1}l_x + c_{3,2}l_y + c_{3,3}m_y + c_{3,0} \quad (16)$$

$$y = c_{4,1}l_x + c_{4,2}l_y + c_{4,3}m_y + c_{4,0} \quad (17)$$

$$m_x = c_{5,1}l_x + c_{5,2}l_y + c_{5,3}m_y + c_{5,0} \quad (18)$$

Where,  $c_{i,y}$  ( $i = 1 \sim 5, j = 0 \sim 3$ ) are all constants.

### 3 Compatible Algebra Equations

In this section, we will give some compatible algebra equations. These equations are very useful to solve the forward kinematics of the parallel mechanisms.

From Eq.(4)~ Eq.(6), we can get:

$$l_z^2 = 1 - l_x^2 - l_y^2 \quad (18)$$

$$m_z^2 = 1 - m_x^2 - m_y^2 \quad (19)$$

$$l_z m_z = -(l_x m_x + l_y m_y) \quad (20)$$

From Eq.(9), Eq.(11), Eq.(12) we can get:

$$z^2 = l_1^2 - x^2 - y^2 \quad (21)$$

$$l_z z = w_1 - l_x x - l_y y \quad (22)$$

$$m_z z = w_2 - m_x x - m_y y \quad (23)$$

Now, from Eq.(18)~Eq.(23), we can get six identities which only contain three unknowns  $l_x, l_y, m_y$ :

$$\begin{aligned} eq_1 : (l_z^2)(z^2) - (l_z z)^2 &= 0 \\ eq_2 : (m_z^2)(z^2) - (m_z z)^2 &= 0 \\ eq_3 : (l_z z)(m_z^2) - (l_z m_z)(m_z z) &= 0 \\ eq_4 : (l_z m_z)(z^2) - (l_z z)(m_z z) &= 0 \\ eq_5 : (m_z z)(l_z^2) - (l_z m_z)(l_z z) &= 0 \\ eq_6 : (l_z^2)(m_z^2) - (l_z m_z)^2 &= 0 \end{aligned} \quad (24)$$



These six identities were presented by Chang-De Zhang and Shin-Min Song in [10]. In addition, we can also find some other information which can be used, if further research work is done carefully.

Let  $\mathbf{W}$  denotes the expression of the position vector  $\mathbf{P}$  in  $O'X'Y'Z'$ , then the relationship between them as follows:

$$\mathbf{W} = \mathbf{R}^T \mathbf{P} \quad (25a)$$

$$\mathbf{P} = \mathbf{R} \mathbf{W} \quad (25b)$$

From Eq.(25), we can see that  $w_1$ 、 $w_2$  are the first two components of  $\mathbf{W}$ , and the following two identities can be obtained:

$$x = l_x w_1 + m_x w_2 + n_x w_3 \quad (26)$$

$$y = l_y w_1 + m_y w_2 + n_y w_3 \quad (27)$$

Eq.(26) and Eq. (27) becomes:

$$n_x w_3 = x - l_x w_1 - m_x w_2 \quad (28)$$

$$n_y w_3 = y - l_y w_1 - m_y w_2 \quad (29)$$

At the same time the lengths of vector  $\mathbf{W}$  and vector  $\mathbf{P}$  are equal, it means:

$$w_3^2 = l_1^2 - w_1^2 - w_2^2 \quad (30)$$

According to the orthogonality and the normalization of the rotation matrix, the following relationships can be obtained:

$$n_x^2 = 1 - l_x^2 - m_x^2 \quad (31)$$

$$n_y^2 = 1 - l_y^2 - m_y^2 \quad (32)$$

$$n_x n_y = -l_x l_y - m_x m_y \quad (32)$$

Now, from Eq.(28)~Eq.(32), we can get another six identities, which only contain three unknowns  $l_x$ 、 $l_y$ 、 $m_y$  :

$$eq_7 : (n_x^2)(w_3^2) - (n_x w_3)^2 = 0$$

$$eq_8 : (n_y^2)(w_3^2) - (n_y w_3)^2 = 0$$

$$eq_9 : (n_x n_y)(w_3^2) - (n_x w_3)(n_y w_3) = 0 \quad (33)$$

$$eq_{10} : (n_x^2)(n_y w_3) - (n_x n_y)(n_x w_3) = 0$$

$$eq_{11} : (n_y^2)(n_x w_3) - (n_x n_y)(n_y w_3) = 0$$

$$eq_{12} : (n_x^2)(n_y^2) - (n_x n_y)^2 = 0$$

These six identities are the additional information. Actually, due to the orthogonality and the normalization of the rotation matrix,  $eq_6$  and  $eq_{12}$  go to the same result. So

there are 11 equations can be used. The information plays an important role in the forward kinematics analysis of parallel mechanism.

### 4 Process of Elimination

In this section, we will formulate 33 equations based on  $eq_1 \sim eq_{11}$ , and then, the method of orthogonal complement are be used to eliminate  $l_x$  and  $l_y$  to obtain a polynomial equation in one unknown  $m_y$ . We multiply the 11 equations  $eq_1 \sim eq_{11}$  by 1、  $l_x$ 、  $l_y$  to form 33 equations with 21 unknown variables.

In order to take advantage of the orthogonal complement method for elimination, the 21 unknown variables can be grouped as follows:

$$\begin{aligned}
 \lambda_1 &= \{1, l_x, l_y, l_y l_x, l_x^2, l_y^2\}^T \\
 \lambda_2 &= \{l_y l_x^2, l_y^2 l_x, l_x^3, l_y^3\}^T \\
 \lambda_3 &= \{l_y l_x^3, l_y^2 l_x^2, l_y^3 l_x, l_x^4, l_y^4\}^T \\
 \lambda_4 &= \{l_y l_x^4, l_y^2 l_x^3, l_y^3 l_x^2, l_y^4 l_x, l_x^5, l_y^5\}^T
 \end{aligned}
 \tag{34}$$

Now there are more equations than the unknown variables, which make the use of the orthogonal complement method for elimination possible. Using the orthogonal complement method the dimension of coefficient matrix can be reduced, and some high order items can be eliminated.

The 33 equations can be simplified as:

$$M_1 \cdot \lambda_1 + M_2 \cdot \lambda_2 + M_3 \cdot \lambda_3 + M_4 \cdot \lambda_4 = \mathbf{0}
 \tag{35}$$

Where  $M_1$  is 4<sup>th</sup> order polynomials of  $m_y$  in theory, and the order of  $M_1$  maybe varies as for the specific problem,  $M_2$  is 2<sup>nd</sup> order polynomials of  $m_y$ ,  $M_3$  are 1<sup>st</sup> order polynomials of  $m_y$ ,  $M_4$  is the constants. If the coefficient matrices are expanded, Eqs.(35) becomes:

$$\begin{aligned}
 &[M_{10} + M_{11}m_y + M_{12}m_y^2 + M_{13}m_y^3 + M_{14}m_y^4]\lambda_1 \\
 &+ [M_{20} + M_{21}m_y + M_{22}m_y^2]\lambda_2 \\
 &+ [M_{30} + M_{31}m_y]\lambda_3 + M_4 \lambda_4 = \mathbf{0}
 \end{aligned}
 \tag{36}$$

The variables  $\lambda_4$ 、  $\lambda_3$ 、  $\lambda_2$ 、  $m_y^3 \lambda_1$  and  $m_y^4 \lambda_1$  can be eliminated by the method of orthogonal complement, and the following equations can be obtained:

$$[N_{10} + N_{11}m_y + N_{12}m_y^2]\lambda_1 = \mathbf{0}
 \tag{37}$$

Denotes  $[N_{10} + N_{11}m_y + N_{12}m_y^2]$  as  $M_s$ , and the order of  $M_s$  maybe varies as for the specific problem. Any 6 rows of  $M_s$  can make up of a square matrix, denotes as  $M$ .

Because Eqs.(37) have non-zero solutions, the determinant of  $\mathbf{M}$  should be equal to zero, so the higher order algebraic equation can be obtained from the determinant of  $\mathbf{M}$ . That is:

$$\det|\mathbf{M}| = 0 \quad (38)$$

From Eq. (38), many polynomial equations of  $m_y$  can be derived, and  $m_y$  can be determined by it. The degree of these algebraic equations may exceed 8. But in most cases, the coefficients of the items of which the order is higher than 8, is very small, so these items can be directly removed. But  $m_y^i (i > 8)$  can be solved, and then substituted into the rest equations. In the end, 8<sup>th</sup> polynomial equations of  $m_y$  can be derived.

As for the other variables,  $l_x$  and  $l_y$  can be uniquely determined by any five equations of Eq.(37). The variables  $w_1$ 、 $w_2$ 、 $x$ 、 $y$ 、 $m_x$  can be determined by Eq.(14)~Eq.(18);  $z$ 、 $l_z$ 、 $m_z$  can be determined by the following three equations:

$$\begin{aligned} z &= \pm\sqrt{l_1^2 - x^2 - y^2} \\ l_z &= \frac{w_1 - l_x x - l_y y}{z} \\ m_z &= \frac{w_2 - m_x x - m_y y}{z} \end{aligned} \quad (39)$$

The last three variables can be determined by the following three equations:

$$\begin{aligned} n_x &= l_y m_z - l_z m_y \\ n_y &= l_z m_x - l_x m_z \\ n_z &= l_x m_y - l_y m_x \end{aligned} \quad (40)$$

Since Eq. (38) is an 8<sup>th</sup> order polynomial equation, so  $m_y$  has 8 possible solutions. From Eq. (39), there are two sets of solutions for each  $m_y$ . Therefore, the forward kinematics of the 6-3 Stewart parallel mechanism has 16 possible solutions.

## 5 Numerical Example

In this section, the algorithm of the forward kinematics of the 6-3 Stewart parallel mechanism is verified through an example. The geometrical parameters of this example are showing as Tab.1.

**Table 1.** The geometrical parameters of 6-3 Stewart parallel mechanism

	$A_x$	$A_y$	$B_x$	$B_y$
1	0	0	0	0
2	0	0	7	0
3	5	0	7	1
4	5	0	4	4
5	2	5	3	4
6	2	5	0	1

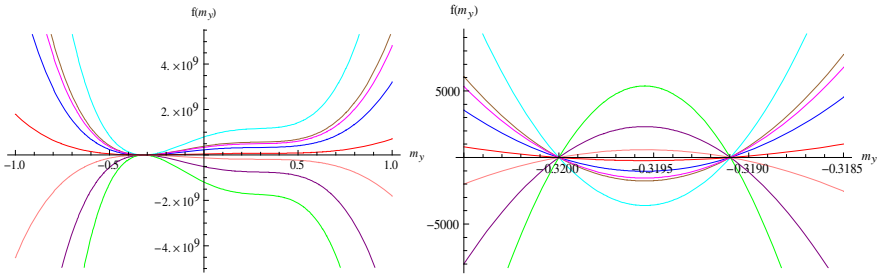
The lengths of the six links are obtained from the inverse kinematics. The position vector is  $\mathbf{P} = [2 \ 3 \ 15]^T$ . The rotation matrix is:

$$\mathbf{R} = \begin{bmatrix} -0.74 & -0.27244 & 0.61496 \\ 0.67244 & -0.32 & 0.667401 \\ 0.0149603 & 0.907401 & 0.42 \end{bmatrix}$$

The lengths of the six links are  $l_1=15.4272$ 、 $l_2=16.0935$ 、 $l_3=18.2124$ 、 $l_4=16.2886$ 、 $l_5=19.98$ 、 $l_6=19.6626$ .

Applying the method proposed in this paper, the forward kinematics analysis can be expressed as 8<sup>th</sup> polynomial algebraic equation, and there are many such equations. Figure 2 shows the curves corresponding to eight such equations. One of them is as follows:

$$\begin{aligned} & -1.18168 \times 10^8 m_y^8 - 1.78846 \times 10^9 m_y^7 + 6.93631 \times 10^8 m_y^6 \\ & - 1.5863 \times 10^9 m_y^5 + 2.26799 \times 10^{10} m_y^4 - 7.71886 \times 10^9 m_y^3 \\ & - 4.36479 \times 10^9 m_y^2 + 2.64473 \times 10^9 m_y + 7.95852 \times 10^8 = 0 \end{aligned}$$



**Fig. 2.** The curves corresponding to eight polynomial equations with respect to the unknown variable  $m_y$ . This shows that there are two solutions between -1 and 1.

These 8<sup>th</sup> polynomial equations have the same solutions, as follows;

$$\begin{aligned} m_y &= -15.6173 & m_y &= -0.319097 \\ m_y &= -0.972551 - 2.209i & m_y &= 0.499764 - 0.317754i \\ m_y &= -0.972551 + 2.209i & m_y &= 0.499764 + 0.317754i \\ m_y &= -0.32 & m_y &= 2.06701 \end{aligned}$$

It is show that there are two solutions between -1 and 1, that is -0.32 and -0.319097. And -0.32 is the initial value, in the other word, the value of  $\mathbf{R}(2,2)$  is -0.32. The calculation result reflects that the proposed method is absolutely correct. The other variables can be obtained easily according to the corresponding equations in this paper, those results are omitted here.

## 6 Conclusions

In this paper, a new method is proposed to solve the forward kinematics problem of the 6-3 Stewart parallel mechanism. Five new compatible equations have been proposed. And the orthogonal complement method has been used to eliminate the redundant variables. This problem can be expressed as an 8<sup>th</sup> polynomial equation in the end, and there are 16 solutions altogether. The method has been verified in a numerical example. All the 16 solutions do satisfy the original six length constraints, that is to say, there is no extraneous solution.

## Acknowledgment

The authors gratefully acknowledge the supported of the National Natural Science Foundation of China (50375071), the supported of Aviation Science Foundation of China (H0608-012), the supported of Jiangsu Province Graduate Research and Innovation Program of China (CX07B-068z).

## References

1. Huang, Z., Kong, L.F., Fang, Y.F.: Mechanism theory and control of parallel manipulator. China Machine Press, Beijing (1997)
2. Liang, C.G., Rong, H.: The Forward Displacement Solution to a Stewart Platform Type Manipulator. *Chinese Journal of Mechanical Engineering* 27, 26–30 (1994)
3. Huang, Z., Zhao, Y.S., Zhao, T.S.: Advanced spatial mechanism. Higher Education Press, Beijing (2006)
4. Nanua, P., Waldron, K.J., Murthy, V.: Direct Kinematic Solution of a Stewart Platform. *IEEE Trans. Robot. Autom.* 6, 438–444 (1990)
5. Akçali, İ.D., Mutlu, H.: A Novel Approach in the Direct Kinematics of Stewart Platform Mechanisms with Planar Platforms. *Asme J. Mech. Des.* 128, 252–263 (2006)
6. Song, S.K., Kwon, D.S.: New Closed-Form Direct Kinematic Solution of the 3-6 Stewart-Gough Platform Using the Tetrahedron Approach. In: *Proceedings of the International Conference on Control, Automation and Systems*, pp. 484–487. ICCAS Press, Jeju (2001)
7. Yurt, S.N., Anli, E., Ozkol, I.: Forward kinematics analysis of the 6-3 SPM by using neural networks. *Meccanica* 42, 187–196 (2007)
8. Ku, D.M.: Forward Kinematic Analysis of a 6-3 Type Stewart Platform Mechanism. *P. I. Mech. Eng. K.-J. Mul.* 214, 233–241 (2000)
9. Yee, C.S., Lim, K.B.: Forward kinematics solution of Stewart platform using neural networks. *Neurocomputing* 16, 333–349 (1997)
10. Zhang, C.D., Song, S.M.: Forward Position Analysis of Nearly General Stewart Platforms. *Asme J. Mech. Des.* 116, 54–60 (1994)

# Portable Multi-axis CNC: A 3-CRU Decoupled Parallel Robotic Manipulator

Dan Zhang<sup>1,2</sup>, Zhen Gao<sup>1</sup>, and Jijie Qian<sup>1</sup>

<sup>1</sup>Faculty of Engineering and Applied Science  
University of Ontario Institute of Technology, Oshawa, ON L1H 7K4, Canada  
Dan.Zhang@uoit.ca

<sup>2</sup>College of Mechanical Engineering  
Dong Hua University, Shanghai, China

**Abstract.** This paper presents a lower mobility parallel manipulator with fully decoupled motions. The proposed parallel manipulator has 3-DOF and can be utilized as a portable multi-axis CNC for parts assembly and light machining tasks that require large workspace and high dexterity. The manipulator consists of a moving platform that is connected to a fixed base by three pairwise orthogonal legs which are comprised of one cylinder, one revolute and one universal joint respectively. The mobility of the manipulator and structure of the inactive joint are analyzed. Kinematics of the manipulator including inverse kinematics, forward kinematics, and velocity equation are deduced. A design optimization is developed to seek the maximal workspace with particle swarm algorithm. This advantage has great potential for machine tools and coordinate measuring machine.

**Keywords:** Parallel manipulator, Decoupled motion, Kinematic modeling, Dynamics simulation, Prototype fabrication.

## 1 Introduction

During the past three decades, parallel robotic manipulators have been the subject of many robotic researches [1-10]. A parallel manipulator typically consists of a moving platform that is connected to a fixed base by at least two kinematic chains in parallel. Parallel manipulators can provide several attractive advantages over their serial counterpart in terms of high stiffness, high accuracy, and low inertia, which enable them to become viable alternatives for wide applications [11-17]. But parallel manipulators also have disadvantages, such as complex forward kinematics, small workspace, complicated structures, and a high cost [18]. To overcome the above shortcomings, progress on the development of parallel manipulators with less than 6-DOF has been accelerated.

Most of the existing parallel robotic manipulators have coupled motion between the position and orientation of the end-effector. The complexity of the multi-axial manipulation produces the difficulty to control. The mechanism with decoupled motion can be called as decoupled mechanism, which has two obvious advantages at least. The first one is that the decoupled mechanism will make the kinematics modeling, Jacobian calculation, and performance analysis becoming more convenient. The second

advantage is the control strategy will be much easier since the displacement, speed, and acceleration of the end-effector in one direction is only driven by one motor that is installed in one limb of the parallel manipulator.

Nevertheless, the number of real applications of decoupled motion actuated parallel manipulators is still quite limited. This is partially because effective development strategies of such types of closed-loop structures are not so obvious. In addition, it is very difficult to design mechanisms with complete decoupling, but it is possible for fewer DOF parallel manipulators. To realize kinematic decoupling, the parallel manipulators are needed to possess special structures. Thus, investigating a parallel manipulator with decoupling motion remains a challenging task.

This paper focuses on the conceptual design, analysis and manufacturing of a 3-DOF non-over constrained translational parallel robotic manipulator with decoupled motions which can be utilized for parts assembly and light machining tasks that require large workspace, high dexterity, high loading capacity, and considerable stiffness. In what follows, conceptual generation, mobility analysis and inactive joint are presented in Section 2. Inverse and forward kinematics modeling is addressed in Section 3. Afterwards, dynamics simulation is conducted based on Lagrangian method in Section 4. In Section 5, particle swarm algorithm based workspace optimization is conducted for the 3-CRU manipulator. In Section 6, the physical prototype is manufactured and tested. Finally, conclusion is given.

## 2 Geometrical Design

A novel 3-CRU (here,  $C$  represents cylindrical joint,  $R$  represents revolute joint and  $U$  represents universal joint) non-over constrained translational parallel manipulator with decoupled motions is proposed. It should be mentioned that 3-CRR mechanism [19, 20, 22] has the similar motion capability. However, this design is an optional solution. As shown in Figs. 1 and 2, the 3-CRU is composed of a base and a moving platform connected by three CRU legs.

The axes of the  $C$  and  $R$  joints, as well as the axe of the  $U$  joint, are arranged such that their joint axes are parallel to a common plane. As a result, the axe of outer  $R$  joint (the one connects to the moving platform) of the  $U$  joint is perpendicular to the axe of the  $U$  joint as well as all the axes of the  $C$  and  $R$  joints within a given leg. All above guarantee that instantaneous rotation of the moving platform about a direction that is perpendicular to the common plane of all the axes of the  $C$ ,  $R$ , and inner  $R$  of the  $U$  joint is impossible.

Since most of the lower-mobility parallel mechanisms are over-constrained, it is necessary to take the common constraints of mechanism and the passive constraints into consideration in mobility analysis. Thus, the DOF of this manipulator can be obtained from the general Chebyshev-Grubler-Kutzbach formula,

$$M = d(\eta - g - 1) + \sum_{i=1}^g f_i + \nu = 6 \times (8 - 9 - 1) + 3 \times (2 + 1 + 2) + 0 = 3 \quad (1)$$

This 3-DOF decoupled parallel manipulator with three CRU legs possesses the following advantages: a) Simple kinematics and easy for analysis, design, trajectory planning, and motion control; b) Large and well shaped workspace; c) High stiffness; d) High loading capacity.

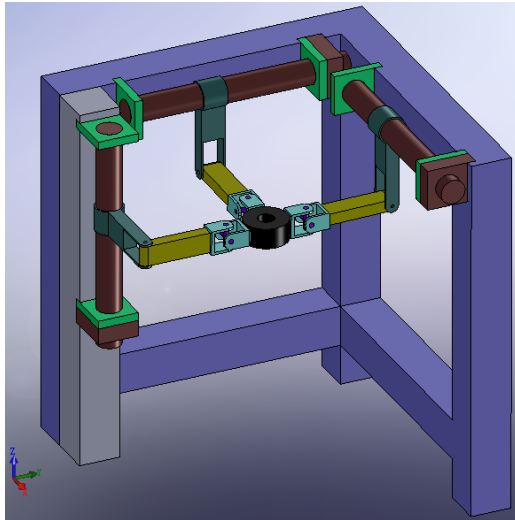


Fig. 1. CAD Model of 3-CRU

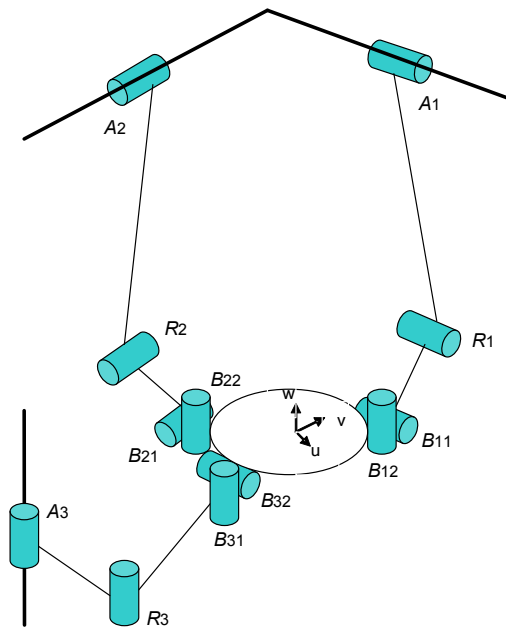


Fig. 2. Schematic model of 3-CRU

A parallel manipulator with less than 6-DOF for each leg can be treated as a twist system that is guaranteed under some exerted structural constraints. The wrench system is a reciprocal screw system of the twist system for the same leg, and a wrench is said to be reciprocal to a twist if the wrench produces no work along the twist.



Concerning the 3-CRU manipulator, the wrench system of a leg is a 1-system which exerts one constraint couple to the moving platform with its axis perpendicular to the axis of C joint within the same leg. The wrench system of the moving platform, that is a linear combination of wrench systems of all the three legs, is a 3-system, because the three wrench 1-systems consist of three couples, which are linearly dependent and form a screw 3-system. Since the arrangement of all the joints shows on Fig. 2, the wrench systems restrict three rotations of the moving platform with respect to the fixed base at any instant, thus leading to a translational parallel manipulator.

### 3 Kinematic Modeling

#### 3.1 Inverse Kinematics

The inverse kinematics problem (IKP) deals with finding the actuated joints' values that correspond to a given set of output variables (position and orientation of the moving platform). The purpose of the inverse kinematics issue is to solve the actuated variables from a given position of the mobile platform.

The IKP of the Stewart-Gough manipulator is trivial with single solution, but when the number of kinematic chains is reduced, the number of solutions of the inverse kinematics problem increases and the problem becomes more challenging [21]. The direct kinematics problem of parallel manipulators is by far more challenging than the inverse kinematics problem since it requires solving a set of polynomial equations in the output variables. While the inverse kinematics problem for a general Stewart-Gough manipulator has only one solution, the direct kinematic problem has up to forty real solutions [22].

To facilitate the analysis, referring to Fig.3, a fixed reference frame  $O-xyz$  is attached at the centered point  $O$  on the base and a moving reference frame  $P-uvw$  at the centered point  $P$  on the moving platform, with the  $z$  and  $w$  axes perpendicular to the platform, and the  $x$  and  $y$  axes parallel to the  $u$  and  $v$  axes, respectively. The direction of the  $i$ th fixed C joint is denoted by unit vector  $\mathbf{c}_i$ . A reference point  $M_i$  is defined on the axis of the  $i$ th fixed C joint, and the sliding distance is defined by  $d_i$ ; point  $B_i$  defined as the interaction of the axes of the U joint. Furthermore, the position vector of point  $M_i$  in the fixed frame is denoted by  $\mathbf{m}_i = [m_{xi}, m_{yi}, m_{zi}]^T$ , while the position vector of point  $B_i$  is noted by  ${}^p \mathbf{b}_i = [b_{xi}, b_{yi}, b_{zi}]^T$  in the moving frame and  $\mathbf{b}_i$  in the fixed frame. The  $\mathbf{u}_i$  is defined as a vector connecting  $A_i$  to  $B_i$ , and is orthogonal to the unit vector  $\mathbf{r}_i = [r_{xi}, r_{yi}, r_{zi}]^T$

Generally, the position and orientation of the moving platform with respect to the fixed base frame can be described by a position vector  $\mathbf{p} = [p_x, p_y, p_z]^T$  and a  $3 \times 3$  matrix  ${}^o \mathbf{R}_p$ .

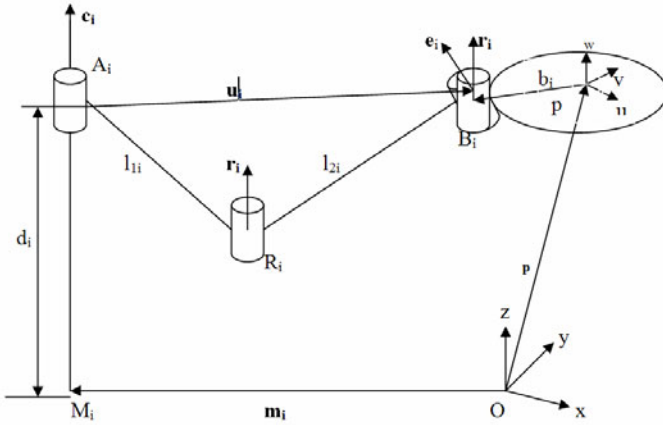


Fig. 3. Kinematic modeling of a leg

Since the moving platform of a 3-CRU possesses only translational motion  ${}^oR_p$  becomes an identity matrix. Then, it has

$$b_i = {}^p b_i \tag{2}$$

Referring to Fig. 3, a vector-loop equation can be written for each leg as follows:

$$u_i = p + b_i - m_i - d_i c_i \tag{3}$$

Substituting Eq. (2) into Eq. (3), it has,

$$u_i = p + {}^p b_i - m_i - d_i c_i \tag{4}$$

As vectors  $u_i$  and  $r_i$  are orthogonal, it yields,

$$u_i^T r_i = 0 \tag{5}$$

Substituting Eq. (4) into Eq. (5), it gets,

$$(p + {}^p b_i - m_i - d_i c_i)^T r_i = 0 \tag{6}$$

For geometric parameters of this parallel manipulator, it has,

$$c_1 = r_1 = [1, 0, 0]^T \tag{7}$$

$$c_2 = r_2 = [0, 1, 0]^T \tag{8}$$

$$c_3 = r_3 = [0, 0, 1]^T \tag{9}$$

$$e_1 = [0, 0, 1]^T \tag{10}$$

$$e_2 = [0, 0, 1]^T \tag{11}$$

$$e_3 = [1, 0, 0]^T \tag{12}$$

Substituting Eqs. (7)~(12) into Eq. (5), the solution of the inverse displacement analysis for the 3-CRU parallel manipulator is derived as,

$$d_1 = p_x + b_{x1} - m_{x1} \quad (13)$$

$$d_2 = p_y + b_{y2} - m_{y2} \quad (14)$$

$$d_3 = p_z + b_{z3} - m_{z3} \quad (15)$$

From Eqs. (13) to (15), it can easily observed that the motions of the 3-CRU parallel manipulator are decoupled. The actuator in leg 1 controls the translation along the X direction; the actuator in leg 2 controls the translation along the Y direction while the actuator in leg 3 controls the translation along Z direction.

The distance between the center of the moving plate and base is  $m_i = l_{2i} + b$

### 3.2 Forward Kinematics

The forward kinematics is to obtain the end-effector position,  $[p_x, p_y, p_z]^T$ , when the input sliding distance  $d_i$  is given.

By solving Eqs. (13) to (15) for variables  $x$ ,  $y$ , and  $z$ , the forward kinematics can be performed. Thus, it yields,

$$p_x = d_1 - b_{x1} + m_{x1} \quad (16)$$

$$p_y = d_2 - b_{y2} + m_{y2} \quad (17)$$

$$p_z = d_3 - b_{z3} + m_{z3} \quad (18)$$

### 3.3 Velocity Analysis

Equations (16) to (18) can be rewritten as:

$$\begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} p_x + b_{x1} - m_{x1} \\ p_y + b_{y2} - m_{y2} \\ p_z + b_{z3} - m_{z3} \end{bmatrix} \quad (19)$$

Taking the derivative of Eq. (19) with respect to time yields,

$$\begin{bmatrix} \dot{d}_1 \\ \dot{d}_2 \\ \dot{d}_3 \end{bmatrix} = \mathbf{J} \begin{bmatrix} \dot{p}_x \\ \dot{p}_y \\ \dot{p}_z \end{bmatrix} \quad (20)$$

where  $\mathbf{J}$  is the  $3 \times 3$  identity matrix. Since  $\mathbf{J}$  is an identity matrix, the manipulator is isotropic everywhere within its workspace.

Therefore, the velocity equations of the 3-CRU parallel manipulator can be written as,

$$\dot{d}_1 = \dot{p}_x, \quad \dot{d}_2 = \dot{p}_y, \quad \dot{d}_3 = \dot{p}_z \quad (21)$$

### 4 Dynamics Simulation

The dynamics analysis of parallel mechanism includes inertial force calculation, stress analysis, dynamic balance, dynamic modeling, computer dynamic simulation, dynamic parameter identification and elastic dynamic analysis, in which dynamic modeling is the most fundamental and important one. Generally, because of the complexity of parallel mechanism, the dynamic model is multi-degree-of-freedom, multi-variable, highly nonlinear and multi-parameter coupling.

There are many approaches for current analysis of dynamics, such as the methods of Lagrangian, Newton-Euler, Gauss and Kane. Lagrangian method can not only seek the simplest form to deduce related dynamics equations of complex system, but also has an explicit structure.

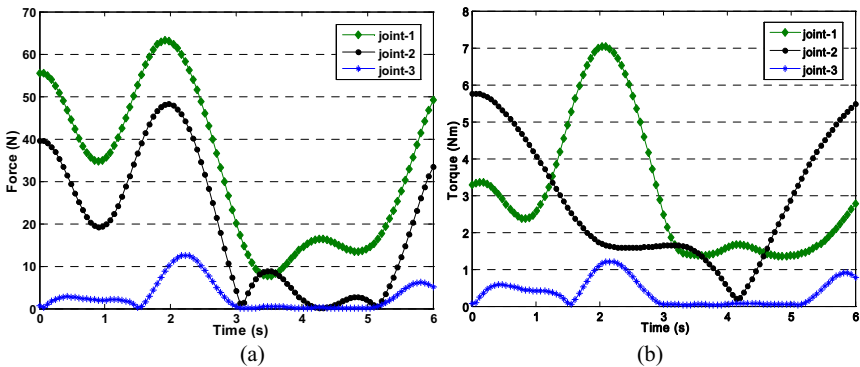
$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_i} \right) - \frac{\partial L}{\partial q_i} = F_i + \sum_{j=1}^k \lambda_j \frac{\partial \Gamma_j}{\partial q_i} \quad (i = 1, 2, 3) \tag{22}$$

where  $L$  is equal to the kinetic energy minus the potential energy of the system,  $q_i$  is the generalized coordinate,  $F_i$  is the generalized force with respect to the generalized coordinate,  $\Gamma_j$  denotes the  $j$ th constraint function,  $k$  is the number of constraint function and  $\lambda_j$  is the Lagrangian multiplier.

The external force applied on the centroid of moving platform is

$$\begin{cases} F_x = 120 \cdot \sin(\text{time}) + 48 \\ F_y = 80 \cdot \cos(\text{time}) + 42 \\ F_z = 45 \cdot \sin(\text{time}) + 45 \cdot \cos(\text{time}) + 48 \end{cases} \tag{23}$$

Under the condition of joints motion and applied force, the curves with time of the driving force and corresponding torque of the four translational joints can be recorded as shown in Fig. 4(a)~(g).



**Fig. 4.** The results of each driven joint (a) force, (b) torque, (c) translational displacement, (d) translational velocity, (e) translational acceleration, (f) angular velocity, (g) angular acceleration

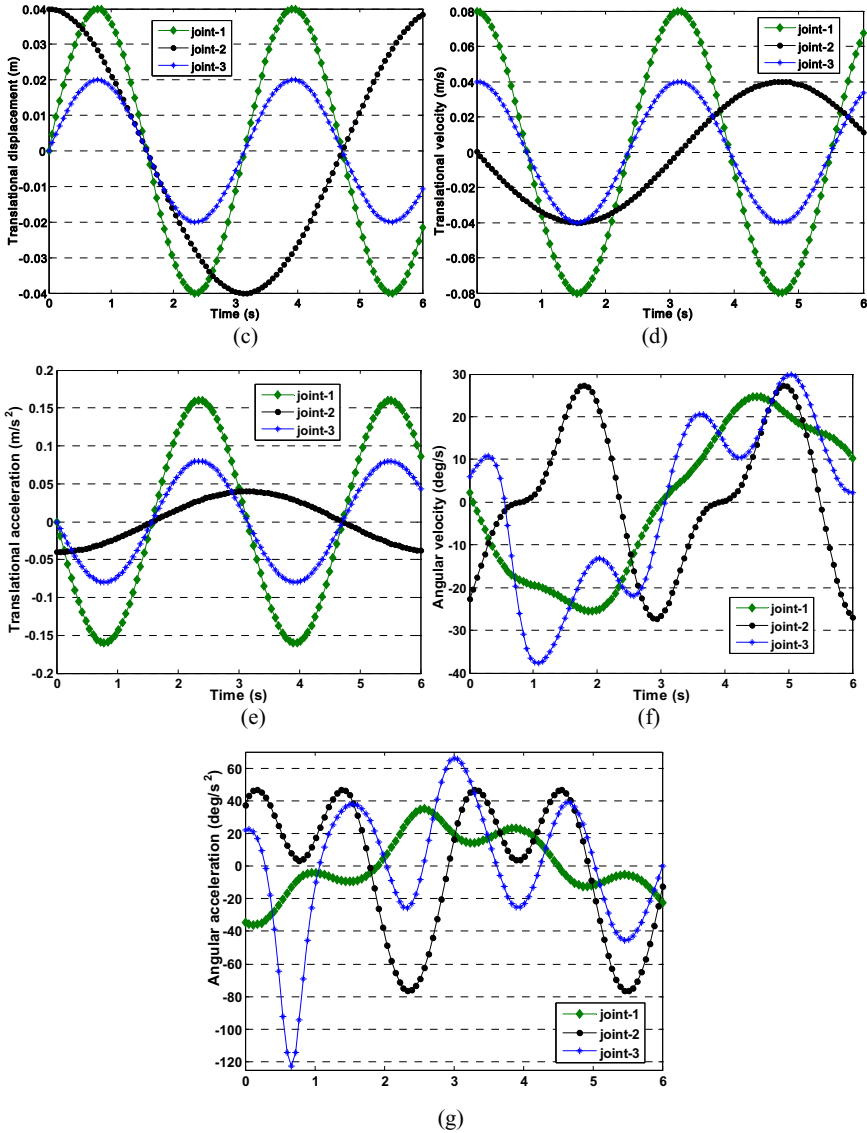


Fig. 4. (continued)

## 5 Workspace Optimization with Particle Swarm Algorithm

The workspace of the parallel robot can be defined as a reachable region of the origin of a coordinate system attached to the center of the moving plate. Since its major drawback is a limited workspace, it is of primary importance to develop algorithms by

which the workspace can be determined and the effect of different designs on the workspace can be evaluated.

Obtaining high performance requires the choice of suitable mechanism dimensions especially as there is much larger variation in the performances of parallel architectures according to the dimensions than for classical serial ones. Indeed, with the development of manipulators for performing a wide range of tasks, the introduction of performance indices or criteria, which are used to characterize the manipulator, has become very important. A number of different optimization criteria for manipulators may be appropriate depending on the resources and general nature of tasks to be performed. The choice of any of the criteria for a given set of data would result in a manipulator whose performances do not necessarily match the optimum values of the other criteria.

Inspired by social behavior such as bird flocking, particle swarm optimization (PSO) is swarm intelligence based stochastic optimization technique. Different with the traditional genetic algorithm, PSO has no evolution operators including crossover and mutation. In PSO, the particles search the optimal solution by following the best particles in the current population. In this scenario, Trelea's model is utilized to perform the optimization process [23]. The maximal velocity divisor is 2, the particles number is 24.

The goal of optimization criteria is to seek the maximal workspace. From Fig.5, it can be found that after 120 generation, the best individuals, namely the optimal workspace, is convergent to  $1.0648 \times 10^7 \text{ mm}^3$ . The computing time is 2.440478 s.

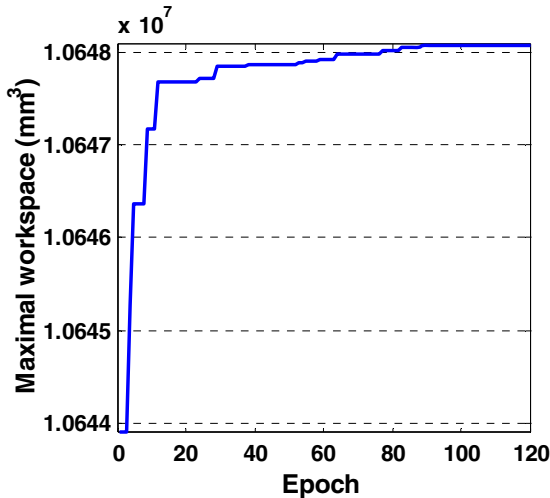


Fig. 5. The best individuals in each epoch with PSO

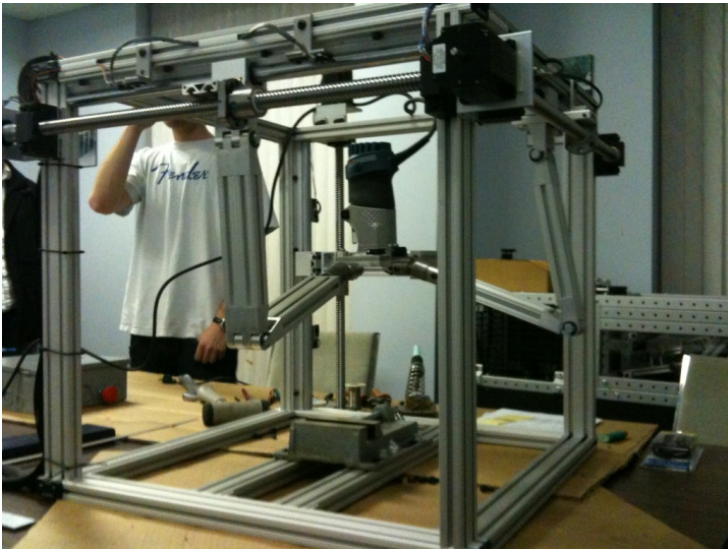
## 6 Prototype and Experiment

To more effectively prove the concept, the prototype is manufactured with some simple materials and basic hand tools. A threaded rod is utilized to create the wanted translations in the three orthogonal directions, by lodging a nut into guide blocks of

the same thread size and pitch. The block is kept tracking straight using a steel guide rod through a hole in the guide block. The threaded rod will act similarly to the ball screws allowing the block to feed in the direction dependant on the screw movement. These threaded rods are powered by drills affixed to nuts brazed onto the end of the rod to crudely simulate the electric stepper motors that will ultimately power our final prototype. The threaded rod in combination with the variable drill speeds represented two types of machine tool motion in the form of a quick rapid to a slow and precise feed. This demonstrates the incredible resolution that can be achieved through this type of motion.

Aside from displaying the motion of the final unit, as well as how the system translates motion through the use of independent linked arms acting on independent axis, but it also displayed the physical size and shape of the overall unit. This includes height, footprint, and the generally robust design that it has. This not only helps the consumer imagine the part size that can be cut, but also how and where the final tool can be located. Figure 6 shows the physical prototype of the 3-CRU decoupled parallel robotic manipulator.

The four characters “UOIT” are selected to test the cutting performance with this portable machine. Beginning with the “U” when the original plunge cut is made, the bit bites into the wood and then applies a force to the arms, that is generally referred to as “walking”. This walking force is the natural force implied on the router as the end mill bites the wood and cuts through it. Due to the amount of play at the end mill, the router simply walks until the slack or play in the arms is taken up, and proceeds straight down as normal (as shown in Fig. 7).



**Fig. 6.** The physical prototype



**Fig. 7.** The characters cut by the portable multi-Axis CNC

## 7 Conclusion

The major contribution of this research is to have proposed a novel 3-DOF parallel manipulator as a portable multi-Axis CNC. The advantage of decoupling in mechanism has great potential for machine tools and coordinate measuring machine. The conceptual design of a novel 3-DOF non-over-constraint translational parallel manipulator with decoupled motion has been proposed. The kinematic modeling of the parallel manipulator has been examined, in terms of inverse and forward kinematics study and velocity analysis. It has been noticed that due to the isotropy and motion decoupling, the inverse and forward kinematic are easy for analysis; The Jacobian is a  $3 \times 3$  identity matrix. The workspace optimization is implemented with particle swarm algorithm. Some of the key issues in the detailed prototype design of the 3-CRU manipulator were discussed as well.

**Acknowledgments.** The authors would like to thank the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC).

## References

1. Wang, L.H., Nace, A.: A Sensor-Driven Approach to Web-based Machining. *J. Intell. Manuf.* 20, 1–14 (2009)
2. Arsenault, M., Gosselin, C.M.: Kinematic, Static and Dynamic Analysis of a Planar 2-DOF Tensegrity Mechanism. *Mech. Mach. Theory* 41, 1072–1089 (2006)
3. Zhang, D., Bi, Z., Li, B.: Design and Kinetostatic Analysis of a New Parallel Manipulator. *Robot. Comput. Integr. Manuf.* 25, 782–791 (2009)
4. Moon, Y.M.: Bio-mimetic Design of Finger Mechanism With Contact Aided Compliant Mechanism. *Mech. Mach. Theory* 42, 600–611 (2007)
5. Palmer, J.A., Dessent, B., Mulling, J.F., Usher, T., Grant, E., Eischen, J.W., Kingon, A.I., Franzon, P.D.: The Design and Characterization of a Novel Piezoelectric Transducer-Based Linear Motor. *IEEE-ASME Trans. Mechatron.* 13, 441–450 (2008)
6. Dong, W., Sun, L.N., Du, Z.J.: Design of a Precision Compliant Parallel Positioner Driven by Dual Piezoelectric Actuators. *Sens. Actuator A. Phys.* 135, 250–256 (2007)
7. Li, Y.M., Xu, Q.S.: A Novel Design and Analysis of a 2-DOF Compliant Parallel Micromanipulator for Nanomanipulation. *IEEE Trans. Autom. Sci. Eng.* 3, 248–253 (2006)



8. Ukidve, C.S., McInroy, J.E., Jafari, F.: Using Redundancy to Optimize Manipulability of Stewart Platforms. *IEEE ASME Trans. Mechatron.* 13, 475–479 (2008)
9. Fang, S.Q., Franitza, D., Torlo, M., Bekes, F., Hiller, M.: Motion Control of a Tendon-based Parallel Manipulator Using Optimal Tension Distribution. *IEEE-ASME Trans. Mechatron.* 9, 561–568 (2004)
10. Mitchell, J.H., Jacob, R., Mika, N.: Optimization of a Spherical Mechanism for a Minimally Invasive Surgical Robot: Theoretical and Experimental Approaches. *IEEE Trans. Biomed. Eng.* 53, 1440–1445 (2006)
11. Zhang, D., Wang, L.: Conceptual Development of an Enhanced Tripod Mechanism for Machine Tool. *Robot. Comput.- Integr. Manuf.* 21, 318–327 (2002)
12. Bi, Z.M., Lang, S.Y.T., Zhang, D., Orban, P.E., Verner, M.: An Integrated Design Toolbox for Tripod-based Parallel Kinematic Machines. *J. Mech. Des.* 129, 799–807 (2007)
13. Chanal, H., Bouzgarrou, B.C., Ray, P.: Stiffness Computation and Identification of Parallel Kinematic Machine Tools. *Journal of Manufacturing Science and Engineering* 131 (2009)
14. Ottaviano, E., Ceccarelli, M., Castelli, G.: Experimental Results of a 3-DOF Parallel Manipulator as an Earthquake Motion Simulator. In: *ASME Conf. Proc. IDETC/CIE* (2004)
15. He, G.P., Lu, Z.: The Research on the Redundant Actuated Parallel Robot with Full Compliant Mechanism. In: *ASME Conf. Proc. MNC* (2007)
16. Lessard, S., Bigras, P., Bonev, I.A.: A New Medical Parallel Robot and its Static Balancing Optimization. *ASME Journal of Medical Devices* 273, 272–278 (2007)
17. Zhang, D., Gao, Z., Song, B., Ge, Y.J.: Configuration Design and Performance Analysis of a Multidimensional Acceleration Sensor Based on 3RRPRR Decoupling Parallel Mechanism. In: *Proc. IEEE Int. Conf. Decision Control*, pp. 8304–8309 (2009)
18. Zhang, D.: *Parallel Robotic Machine Tools*. Springer, New York (2009)
19. Kong, X.W., Gosselin, C.M.: *Type Synthesis of Parallel Mechanisms*. Springer, New York (2007)
20. Gosselin, C.M., Masouleh, M.T., Duchaine, V., Richard, P.L., Foucault, S., Kong, X.W.: Parallel Mechanisms of the Multiparteron Family: Kinematic Architectures and Benchmarking. In: *IEEE International Conference on Robotics and Automation*, pp. 555–560 (2007)
21. Merlet, J.P.: *Parallel Robots*, 2nd edn. Springer, Heidelberg (2006)
22. Tsai, L.W.: *Robot Analysis: the Mechanics of Serial and Parallel Manipulators*. Wiley, New York (1999)
23. Trelea, I.C.: The Particle Swarm Optimization Algorithm: Convergence Analysis and Parameter Selection. *Information Processing Letters* 85, 317–325 (2003)

# Carton Motion-Moment Diagram and Stiffness Characteristics

Guowu Wei, Ruirui Zhang, and Jian S. Dai\*

School of Physical Sciences and Engineering, King's College London,  
WC2R 2LS London, United Kingdom

guowu.wei@kcl.ac.uk, ruiruizhang527@gmail.com,  
jian.dai@kcl.ac.uk

<http://www.kcl.ac.uk/schools/pse/diveng/research/cmms/jsd/>

**Abstract.** To achieve the versatility in developing a packaging machine, it is imperative to understand the motion and stiffness characteristics of cartons during industrial folding and manipulation. It is desirable to describe carton and its folding procedure in a way that is intelligible by a programmable machine. This raises much interest in the study of carton folding and carton motion for packaging industry. This paper focuses on the stiffness characteristics of carton creases and panels, investigates the mechanics of creases and panels during carton folding and manipulation and characterizes the motion-moment of carton creases during folding. Experiments are carried out to investigate the characteristics of carton creases and panels.

**Keywords:** packaging carton, motion-moment diagram, stiffness, equivalent mechanism.

## 1 Introduction

Diversity, complexity and short runs of packaging present a problem in a mechanized production to the point that manual effort is still resorted to. Examples of such packaging are commonly found in the departmental stores ranging from food to the luxury end of the personal products including cosmetics and perfumery. These involve a very wide range of sizes and variety of shapes of cartons and other packaging materials with frequent innovation and changing characteristics of prestigious products. This can be seen in various seasonal products during Christmas and Easter and induces two distinct actions. One places the onus on the package designer [1] to constrain package variety and novelty to be within the capability of existing commonly preferred machines. The other places the onus on a machine designer to design an automatic and reconfigurable system for all eventualities. The latter is one of the fundamental issues in packaging machine innovation.

Current machines do not have the ability to manipulate the entire range of cartons that either exist or may materialize through innovation. This is reflected in carton erectors/fillers that are normally designed to accept a range of carton sizes, but few

---

\* Corresponding author.

variations of shape and style [2]. This is the reason that a large number of manual operations are still used in packaging.

The academic interest in carton folding and manipulation was only aroused in recent years. Dai and Rees Jones [3] developed a new approach and equated a cardboard carton with a mechanism that generated a new class of mechanisms. In such equivalence, the mechanism analysis including the topological graph and adjacency matrix approach [4] can be used to describe a carton that kinematics of a carton can be identified [5]. The analogy characterizes a carton folding procedure with detailed steps and describes folding in an analytical way. In 1998, the concept of the metamorphic mechanism [6] was proposed relating a carton fold to a new type of mechanisms which change the mobility and mechanism structure during motion and subsequently change the manipulation configurations [7, 8]. In 2000, Lu and Aketa developed a new algorithm [9] for folding a carton with fixtures. The fixtures were arranged to fold different cartons and motion planning was used for folding operation. In 2001, Liu and Dai proposed a method [4] for extracting carton folding sequence and developed an algorithm [10] to automatically fold a carton. In 2002, Dai and Rees Jones developed kinematic models [11] for carton folds and applied kinematic transformation [12] to carton manipulation.

On the other hand, to develop a versatile packaging machine, it is imperative to understand motion and stiffness characteristics of cartons in folding and manipulation. In this part development, Carlsson et al. [13] investigated the residual moment using the finite-element method and examined local bending of a carton panel. Song and Amato [14] used the correlation between box folding and protein folding to investigate paper craft folding by introducing a tree-like structure to identify the links between panels. In parallel to this part of study, Ashdown [15] applied finite-element analysis to examine the apple packaging. Gilchrist et al. [16] used the non-linear finite element to model corrugate boards. Beldie et al. [17, 18] investigated paperboard packages using finite-element modelling. Hicks et al. [19] used a finite-element approach and an energy formulation method to model the delaminating buckling in a geometrically constrained system and compared the results with the experimental results. The minimum energy principles were further used to create a theoretical model [20] of the response of the packaging material during processing by encapsulating the non-linear properties of the material in parametric models generated through analysis of the physical measurement.

Although the various investigations were made, the residual moment produced from a crease of a carton needs to be investigated and is essential in investigating variable stiffness of carton creases and panels. Thus, in this paper test rig is set up in order to test and characterize the stiffness of carton creases and panels. The study firstly investigates the mechanics of creases during carton folding and manipulation and characterizes the motion moment of carton creases during folding. Then stiffness model of the carton panel is established based on mechanics of material and experiment for testing the stiffness of carton panel is carried out resulting in the stiffness characteristics of the panel specimen. This paper aims at developing an analytical way of modelling carton creases with their variable stiffness. The study gives an experimental way of studying and modelling a carton in folding and manipulation in packaging industry, and leads to carton-packaging automation with predictable characteristics.

## 2 Cartons and Their Equivalent Mechanisms

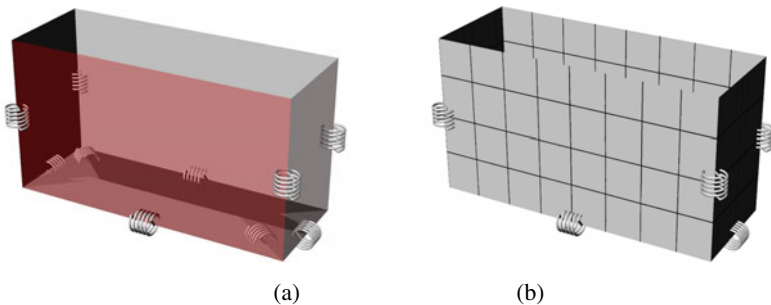
A crash-lock carton presented in Fig. 1 is selected as an example for this study. The bottom of this rectangular carton is closed by interlocking two main based panels each of which is linked to a triangle panel with a crease and to the side panels of the main body with several other creases.

The carton is erected and folded from a flattened configuration by closing the crash-lock base. This process can be modelled as an equivalent mechanism. Consider creases between panels as revolute joints [2, 13] and carton panels as mechanism links [9, 14], the carton can be modelled as a mechanism. This equivalent mechanism can be carried out by assuming carton panels as rigid links when carton panels are made of rigid cardboards. However consider a deformation resulting from the carton folding and manipulation, and particularly creases which have a non-linear stiffness, the equivalent mechanism is represented in Fig.2a where links are rigid. With non-rigid panels it has been modelled in Fig. 2b.



**Fig. 1.** A crash-lock carton

The mechanism has its particular characteristics that the joint possesses a non-linear stiffness and the link is non-rigid. The study of the carton equivalent hence involves the study of these two characteristics: stiffness of carton creases and stiffness of carton panels. Therefore, using the experiment rig set up in this research, experiments of testing the stiffness of carton creases and carton panels are carried out and the stiffness of carton creases and panel is characterized in this paper.



**Fig. 2.** Equivalent mechanism with nonlinear spring and rigid/non-rigid panels

### 3 Stiffness Characteristics of Creases

In this section, experiment was carried out to test the stiffness characteristics of creases of a carton. The thickness of cardboard panels was 1mm and they are composed of six flies of sheets. The crease was formed by die-pressed groove. Eight creases were labeled in the carton as in Fig. 3.

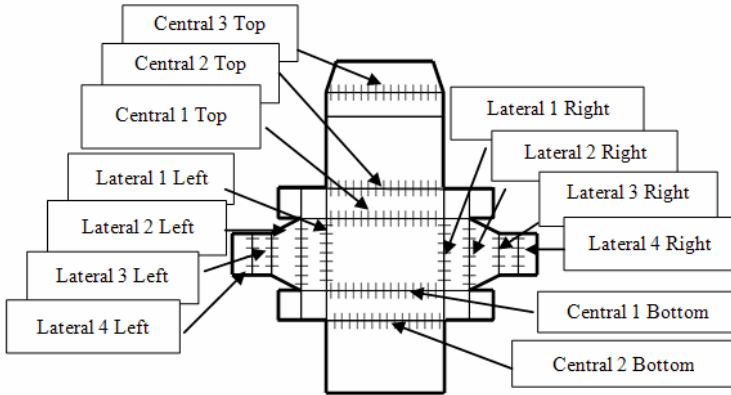


Fig. 3. A test carton

The test carton is fixed on a carton rotating bearing in Fig. 4 which permits a carton to be manipulated about the crease axis. The alignment of the crease axis with the rotational joint axis of the measurement chain is fundamental to avoid an undesired input in measurement, e.g. extra stress and strain of a cardboard carton panel.

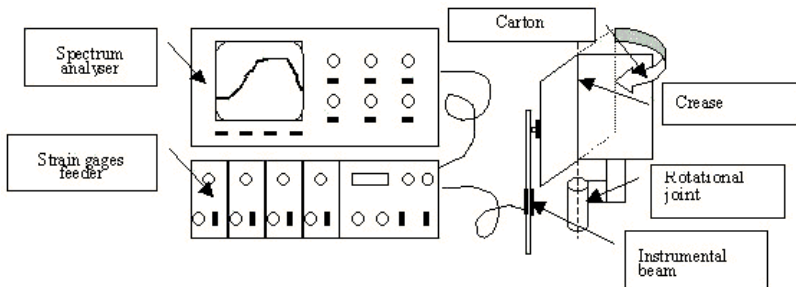


Fig. 4. A measurement chain

The test condition was kept in the experiment to follow the procedural standard [21, 22]. In particular, the laboratory was kept at  $23^{\circ}\text{C} \pm 1^{\circ}\text{C}$  in temperature and  $(50 \pm 2)\%$  in relative humidity at least for 3 hours prior to the start of an experiment.

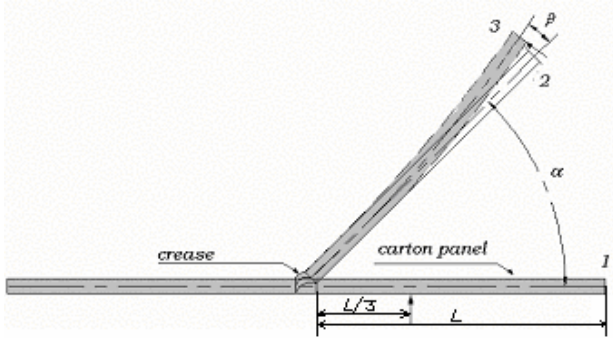


Fig. 5. Effect when rotating a panel about a crease

The carton folding angle about a crease is as in Fig. 5 with a deflection angle of the panel flexibility. This deflection angle is

$$\frac{d}{dx} y(x) = \beta(x) = \frac{FL}{EJ} x - \frac{Fx^2}{2EJ}, \tag{1}$$

where  $x$  is the length coordinate,  $F$  is the applied force and,  $L$  is the length of the panel,  $E$  the Young’s Modulus and  $J$  the moment of inertia. When the manipulating finger is exerted at the  $1/3$  length of the panel, the deflection angle is

$$\beta\left(x = \frac{L}{3}\right) = \frac{FL^2}{18EJ}. \tag{2}$$

### 3.1 Motion-Moment Characteristics

When being exerted a force, a cardboard panel folds about a crease towards another panel to change the folding angle. This force results in a moment exerting to the crease. With the experiment measurement chain both panel manipulation and moment applied to the crease were recorded in Fig. 6.

It can be noted that the data is quite converging. A clear trend is shown in Fig. 6. The upper trend is the first manipulation about creases and the lower trend is the repeat manipulation about creases when plies were detached after the first manipulation. The crease behaviour is similar in each part of a carton.

In the first manipulation, in the stage before  $25^\circ$ , the plies are still attached to each other and give a linear behaviour to the crease stiffness. After this point the moment is linearly proportional to the folding angle and then the stiffness is constant until folding angle reaches  $50^\circ$ . Around  $50^\circ$  the plies begin to detach from each other. In the range between  $50^\circ$  and  $65^\circ$ , most part of the crease cross section is delaminated. The stiffness decreases immediately and becomes several times smaller. In this range the stiffness reaches the lowest value. When the panel is manipulated close to  $90^\circ$ , these detached plies are heaped. They start working against the continuing manipulation about this point and act as a wedge. The crease stiffness starts increase proportionally to the ply compression from this point. The repeat manipulation has much lower stiffness but after the folding angle reaches the final stage at  $90^\circ$ , the crease has the same stiffness. All three manipulations have the converged trends and the stiffness at the final stage reaches the same value.

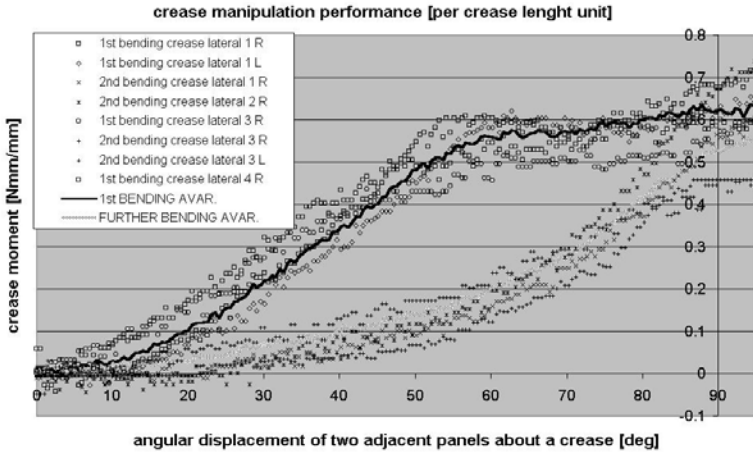


Fig. 6. Experimental motion-moment diagram

This demonstrates four ranges in the motion-moment diagram. They are elastic range, delaminating (yield) range, low panel resistance range and rebounding range. Same characteristics can be found in other creases as in lateral 2, lateral 3, lateral 4 and central panels in Fig. 3. Each has the measurement data of the 1st manipulation and repeat manipulation.

Extracting from the above experimental data, the folding analysis of three stages is given in Fig. 7.

In Fig. 7, crease stiffness is calculated and the three stages of motion and moment characteristics are produced. It is expected that the crease moment increases as a parabolic curve during the first stage, due to the reason that the fibers are stretched until ply delamination. The second stage results in ply delamination. The plies begin to detach each other [23] and the moment required to manipulate cardboard panels about the creases begins to decrease. When these delaminated layers are compressed there is the third stage when the moment is required for the folding and manipulation.

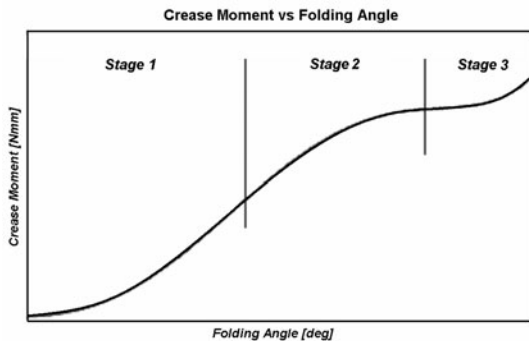
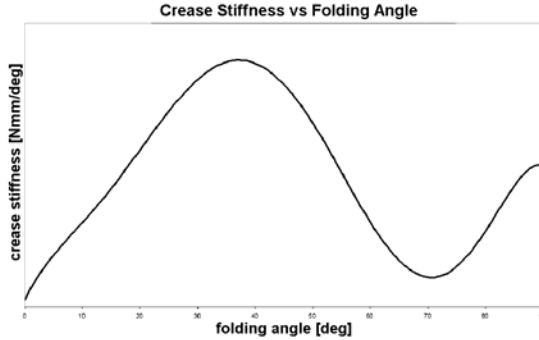


Fig. 7. Crease motion-moment diagram

### 3.2 Stiffness Characteristics of Carton Creases as Equivalent Elastic Joints

From the motion-moment diagram of creases, crease stiffness can be extracted. The stiffness characteristics of creases are hence demonstrated in Fig. 8.



**Fig. 8.** Crease stiffness per length unit vs folding angle

The elastic stage starts from the origin until the yield point before the folding angle reaches  $37^\circ$  where delamination begins to decrease the crease stiffness. In this interval the stiffness increases linearly. The plastic stage starts when the folding incurs the change of mechanical property of a crease. The plies begin to detach from each other. In the range between  $37^\circ$  and  $70^\circ$ , most part of the crease cross section is delaminated. The stiffness decreases immediately. Around  $70^\circ$ , the stiffness reaches the lowest value. The third stage starts when plies are completely detached and cannot resist folding. However, the detached plies are heaped and started working against the continuing manipulation. This stage ends when the layers begin to be compressed by further folding.

Averaged crease stiffness value of different panels for each ranges are shown in Table 1.

**Table 1.** Crease stiffness from several creases

CREASE	Elastic range (stage 1)		Delamination (yield) range (stage 2)		Rebounding range (stage 3)	
	Angle [deg]	Stiffness [Nmm/deg]	Angle [deg]	Stiffness [Nmm/deg]	Angle [deg]	Stiffness [Nmm/deg]
CD 1	0-35	0.01410	35-66.	0.01273	66-90	0.00370
CD 2	0-34	0.01221	32-66	0.01633	66-90	0.00290
CD 3	0-33	0.01013	33-66	0.01559	66-90	0.00400
...	...	...	...	...	...	...
Avarage	0-34	0.01215	34-66.3	0.01488	66-90	0.00351
stand dev	3.3%	16.3%	0.62%	12.8%	4.54%	15.8%



The measurement is accurate to 1.5%.

It should be pointed out that in Table 1 the average and standard deviation are computed using the following formulas [24]:

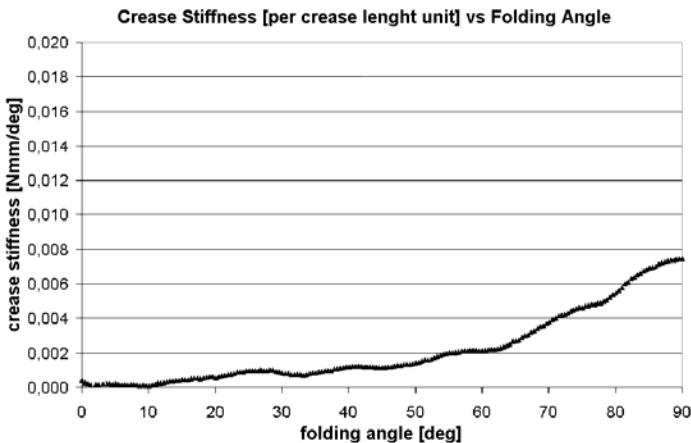
$$\bar{x} = \frac{\sum_1^n x_n}{N}, \quad (3)$$

$$\sigma = \sqrt{\frac{\sum_1^n (x_i - \bar{x})^2}{N}}. \quad (4)$$

### 3.3 Stiffness Characteristics of Carton Creases in Repeat Folding

The repeat manipulation has much lower stiffness as in Fig. 9 but after the folding angle reaches the final stage at 90°, the crease has the same stiffness. All three repeat manipulations have the converged trends and the stiffness at the final stage reaches the same value.

It is interesting to note that the stiffness of the first manipulation and that of the repeat manipulation begin to reach the same value after 70°-80° in folding. This is because after the delamination both manipulations only result in the compression of detached plies. A strong indication is that both curves have the same final stage with a similar folding range when the plies begin to be compressed at the same folding angle.



**Fig. 9.** Crease stiffness per length unit vs folding angle in the repeat manipulation

## 4 Stiffness Characteristics of Carton Panels

### 4.1 Stiffness Model of Carton Panels

A carton panel is formed by a cardboard with a relative high stiffness. The cross-sectional area of the panel is  $wt$ , where  $w$  is the board width and  $t$  is the thickness. The moment of inertia is

$$J = \frac{wt^3}{12}, \tag{5}$$

and the flexural rigidity is  $EI$ . For a small value of deflecting angle  $\beta$ , the bending moment which is applied to a carton panel is related to the carton panel stiffness by a linear relation as

$$M = k_c \beta. \tag{6}$$

If  $x$  is the deflection of a panel, the moment  $M$  can be given as

$$M = 2 \frac{EJ}{L} \frac{x}{L}. \tag{7}$$

For a small angle, the carton panel stiffness is

$$k_p = \frac{EJ}{L}, \tag{8}$$

where the  $x/L$  is the deflection for small angle.

### 4.2 Test of Carton Panel Stiffness

Carton panel is a multiplied cardboard with a much stronger stiffness than that of a crease. The panel is modelled as a set of high stiffness springs to reflect the panel deflection during manipulation. This requires identifying the Young modulus and bending stiffness of a panel by using a tensile test and a resonance test.

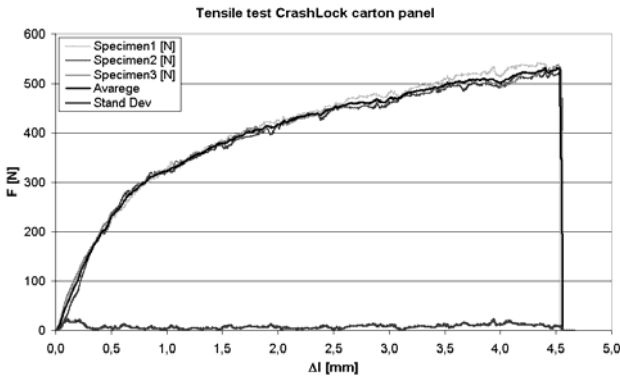


Fig. 10. Force versus elongation of a carton panel

Carton panel tensile test is similar to the metal test. A strip of panel, typically 25mm wide, was clamped between two jaws of a tensile tester. An upper jaw was moved by a lead screw, which gave a constant rate of elongation, stretched the specimen. Load cell and dial gauge were provided to measure force and displacement. A graph of the tensile force/stretch relationship was then drawn. The shape of this curve results from the progressive breaking of inter-fiber bonds plus some breakage of individual fibers.

Constant rate of elongation method were used for calculating the tensile index, the tensile energy absorption index and the modulus of elasticity. The results are shown in Fig. 10.

Three specimens were used. Good agreement is evidenced by the standard deviation, which is less than 7%.

To determine Young Modulus, following expression is used:

$$E = \frac{\sigma}{\varepsilon} = \frac{Fl_0}{S(l_f - l_0)} \quad (9)$$

where  $S$  is the cross section,  $l_0$ ,  $l_f$  is the initial and final lengths of specimen and  $\varepsilon=0.2\%$  is chosen. Young's modulus from the experiment is  $E=2182(\text{N}/\text{mm}^2) \pm 5.71\%$ .

The inertia moment is

$$J = \frac{bt^3}{12} = 0.7146\text{mm}^4, \quad l = 100\text{mm} \quad (10)$$

Hence panel specimen stiffness from Eq. (8) is  $k_p=6.81\text{Nmm}$ .

## 5 Error Analysis

The standard deviation is used to check the convergence of a number of tests. The general analysis has shown a good agreement in the test and in the experiment. In particular, the trend of the standard deviation from the experiment on the motion-moment diagram as in Fig. 6 is produced in Fig. 11.

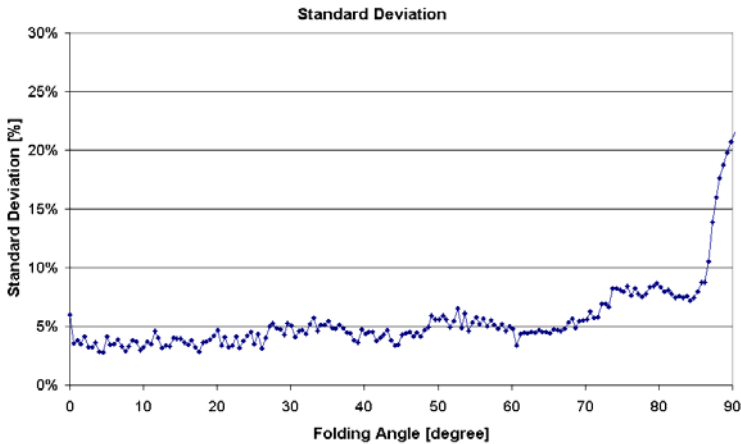


Fig. 11. Standard deviation in measurements of manipulating motion and moment

With the data from the experiments of all crease lines in the range of folding angle between  $0^\circ$  and  $80^\circ$ , the average standard is below 5%. Between  $75^\circ$  and  $87^\circ$ , the standard deviation is below 8%. Hence the characteristics of creases are coherent and the analysis gives a good agreement among the results.

## 6 Conclusions

The investigation of stiffness characteristics of a carton was demonstrated in this paper. The study started from a physical model of creases, revealed the behaviour and variable stiffness of creases and produced a motion-moment diagram for creases. The crease stiffness was analysed as being variable with the folding angle. The investigation of panel stiffness was then progressed. A physical model of carton panel was established revealing the behaviour and variable stiffness of panel. The investigation of experiments generated the Young's modulus and characterized stiffness of carton panel. The error analysis was presented with the illustration of standard deviation. This study provided an experimental way of studying and modelling a carton in folding and manipulation in packaging industry, and led to carton-packaging automation with predictable characteristics.

## References

1. Soroka, W.: Fundamentals of Packaging Technology. In: Emblem, A., Emblem, H. (eds.) Institution of Packaging (1996)
2. Stewart, B.: Packaging as an Effective Marketing Tool. Kogan Page Ltd., London (1996)
3. Dai, J.S., Rees Jones, J.: Mobility in Metamorphic Mechanisms of Foldable/Erectable Kinds. ASME Trans. J. Mech. Des. 121(3), 375–382 (1999)
4. Liu, H., Dai, J.S.: Carton Manipulation Analysis Using Configuration Transformation. Proc. Instn. Mech. Engrs., Part C: J. Mechanical Engineering Science 216(C5), 543–555 (2002)
5. Dai, J.S., Rees Jones, J.: Kinematics and mobility Analysis of Carton Folds in Packing Manipulation. Proc. Instn. Mech. Engrs., Part C: J. Mechanical Engineering Science 216(C10), 959–970 (2002)
6. Dai, J.S., Rees Jones, J.: Mobility in Metamorphic Mechanisms of Foldable/Erectable Kinds. In: 25<sup>th</sup> ASME Biennial Mechanisms and Robotics Conference, DETC98/MECH5902 (1998)
7. Dai, J.S., Rees Jones, J.: Configuration Transformations in Metamorphic Mechanisms of Foldable/Erectable Kinds. In: Proc. of 10th World Congress on the Theory of Machines and Mechanisms, Oulu, Finland, vol. 2, pp. 542–547 (June 1999)
8. Yao, W., Dai, J.S.: Dexterous Manipulation of Origami Cartons With Robotic Fingers Based on the Interactive Configuration Space. Transactions of the ASME: Journal of Mechanical Design 130(2), 223031–223038 (2008)
9. Lu, L., Akella, S.: Folding Cartons with Fixtures: A Motion Planning Approach. In: IEEE International Conf. on Robotics and Automation, Detroit, MI, pp. 1570–1576 (1999)
10. Liu, H., Dai, J.S.: An Approach to Carton-Folding Trajectory Planning Using Dual Robotic Fingers. Robotics and Autonomous Systems 42(1), 47–63 (2003)
11. Dai, J.S., Rees Jones, J.: Kinematics and Mobility Analysis of Carton Folds in Packing Manipulation. Journal of Mechanical Engineering Science, Proc., IMechE. 216(C10), 959–970 (2003)

12. Dai, J.S., Rees Jones, J.: Matrix Representation of Topological Configuration Transformation of Metamorphic Mechanisms. *ASME Transactions, Journal of Mechanical Design* 127, 837–840 (2005)
13. Carlsson, L., Fellers, C., Westerlind, B.: Finite Element Analysis of the Creasing and Bending of Paper. *Svensk Papperstidning* 15, 121–126 (1982)
14. Song, G., Amato, N.M.: A Motion Planning Approach to Folding: From Paper Craft to Protein Folding. In: *IEEE International Conference on Robotics and Automation*, Seoul, pp. 948–953 (2001)
15. Ashdown, K.G.: Finite Element Analysis of Apple Packaging. Master Dissertation Massey University, New Zealand (1995)
16. Gilchrist, A.C., Suhling, J.C., Urbanik, T.J.: Nonlinear Finite Element Modeling of Corrugate Board. *ASME, Mech. Cellulosic Mater. AMD-231/MD-85*, 101–106 (1999)
17. Beldie, L., Sandberg, G., Sandberg, L.: Paperboard Packages Exposed to Static Load – Finite Element Modeling And Experiments. *Packaging Technol. Sci.* 14, 171–178 (2001)
18. Beldie, L., Sandberg, G., Sandberg, L.: Finite Element Modelling and Experiments on Paperboard Packages with Emphasis on Creases. *Packaging Technol. Sci.* (2001)
19. Hicks, B.J., Mullineux, G., Berry, C., McPherson, C.J., Medland, A.J.: Energy Method for Modeling Delamination Buckling in Geometrically Constrained System. *Proc. Instn. Mech. Engrs., Part C: J. Mechanical Engineering Science* 217(C9), 1015–1026 (2003)
20. Hicks, B.J., Berry, C., Mullineux, G., McPherson, C.J., Medland, A.J.: An Energy-Based Approach for Modelling the Behaviour of Packaging Material during Processing. *Proc. Instn Mech. Engrs., Part C: J. Mechanical Engineering Science* 218(C1), 105–118 (2004)
21. Markström, H.: The Elastic Properties of Paper – Test Methods and Measurement Instruments. *Lorentzen & Wrette*, Stockholm (1993)
22. Fellers, C., Carlsson, L.: Measuring the Pure Bending Properties of Paper. *Tappi Journal* 62(8), 107–109 (1979)
23. Guyot, C., Bacquet, G., Schwob, J.M.: Folding Resistance of Magazine Papers. In: *Tappi Proceedings of 1992 Coating Conference*, pp. 255–268 (1992)
24. Doebelin, E.O.: *Measurement Systems Application and Design*. McGraw-Hill, Singapore (1990)

# Flexible Foot Design for Biped Walking on Uneven Terrain

Chen Koon Toh, Ming Xie, Hejin Yang, Guoqing Zhang,  
Quoc Phuong Bui, and Bo Tian

School of Mechanical and Aerospace Engineering, Nanyang Technological University,  
50 Nanyang Drive, Research Techno Plaza, Level 4, BorderX Block, Singapore 637553  
{cktoh, mmxie, hjyang, zhanggq, qpbui, tianbo}@ntu.edu.sg

**Abstract.** Mechanical design of feet for humanoid robot is described to be used with planning and control strategy to overcome issues on uneven terrain walking. The gait planning and control method are explained briefly in this paper, while the foot mechanism design to complement the control algorithm is presented with full details. The design limitations and specifications are reported and the design criteria cum selection of mechanical components are closely examined. Important functions like obstacle detection and shock absorption are included in the design with low cost approach. The 3D drawing of final assembly and finished prototype are included.

**Keywords:** Humanoid Robot, Flexible Foot, Biped Walking, Uneven Terrain.

## 1 Introduction

Humanoid robots nowadays are able to perform various human-like movements like walking, running, turning and stairs climbing. However, most of the movements are achieved in laboratory environment, with flat and smooth floor. In order to utilize humanoid robots in assisting our daily jobs, they must be able to move smoothly in human's living environment, which include inclined slope and rough terrain like tarred road.

There are some interesting approaches in designing of flexible humanoid robotic feet for greater efficiency and safety in terms of encountering uneven surface and slope walking. Basically, they can be divided into four sub-categories: Tactile sensing [1, 2], multi degree of freedom [3, 4, 5, 6] and multi-sensory feedback [7] and shock absorbing compliance [8].

The tactile sensing method is useful in collecting the ground information, for example slope and irregularity by using a few pressure sensors [1] or constant scanning of a pressure sensor grid [2]. The peak pressure values obtained from the readings indicate the contact point of foot with ground. An accurate value of centre of pressure can be calculated from the measurement and used in the control logic.

For foot design with multi degree of freedom, it aims to provide better conformance to the irregular shape of uneven terrain to provide better grip and stability for the robot. Design from Yang utilized flexible cable to change the shape of the foot by manipulating the orientation of four metal plates that formed the foot of robot [3],

while Hashimoto designed an interesting robotic feet that have cam-type locking mechanism which assures the robot to maintain four-point contact on the ground while walking [5, 6]. While Yang's design involves the shape changing of whole feet, Hashimoto's foot needs to change the height of four contact legs only to adapt to the shape of the ground. Besides that, an arc-shape foot is a relatively simpler mechanism from Minakata to give compliance to the shape of uneven terrain [4].

In order to achieve more accuracy in decision making for uneven terrain walking, multiple sensors like MEMS accelerometer and angular rate gyroscopes can be installed to foot of biped robot [7]. Other than that, multiple layers of shock absorbing rubber pads are sandwiched between top and bottom plates of WAF-2 to permit a small angle of inclination between the foot structures itself [8].

Other than that, there is another approach to uneven ground walking by using tools. HOAP-3 from Tokyo Denki University was demonstrated to walk across uneven terrain with the help of a stick [9].

In this paper, the design of low cost foot mechanism to assist the control algorithm planned for uneven terrain walking. The remainder of paper is organized as follows. Section 2 presents the overview of hardware system and software architecture built. Next, section 3 discusses the development of foot design with respect to specified criteria. In section 4, detail hardware design is delivered. Lastly, conclusion is addressed in section 5.

## 2 Background

Low cost humanoid robot (LOCH) is developed for study in human-assisted manipulation and biped locomotion of humanoid robot. The hardware and software systems have been built and tested on flat ground.

### 2.1 Hardware Systems

LOCH robot has 50 degrees of freedom (DOF) in total, which is 12 DOFs at its two arms, 20 DOFs at two hands, 2 DOFs for torso, 2 DOFs at its neck and 14 DOFs for

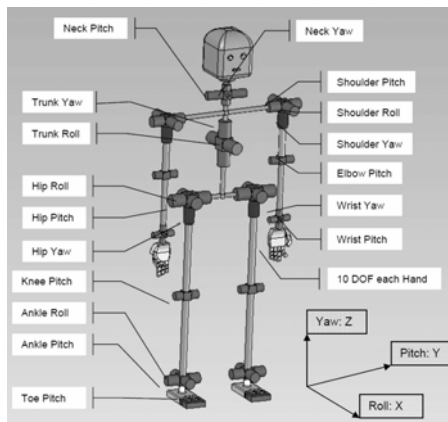


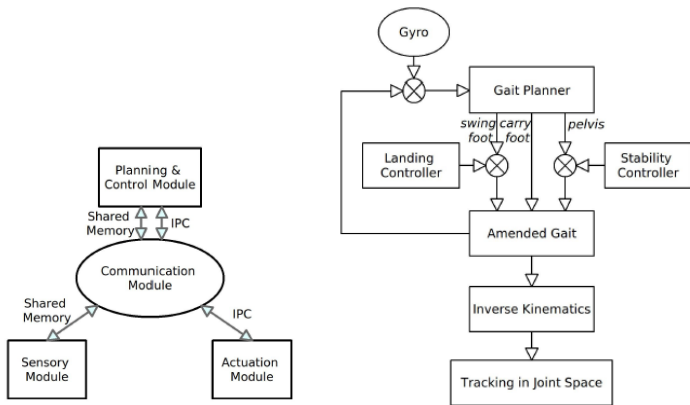
Fig. 1. DOFs layout of LOCH robot

biped systems. The respective degree of freedom can be referred from Fig. 1. All major joints are using timing belt-harmonic drive system as explained in [10].

While for the electronic hardware system, it comprised of a PC104, four 6 axis force-torque sensors (installed at wrists and ankles), a gyroscope, 14 motor controllers and CAN bus as the main communication system. The combination of the sensors provide required data about the motion status and floor conditions to evaluate the stability during walking process utilizing zero moment point (ZMP) concept.

### 2.2 Software Architecture

The software architecture of LOCH robot is designed to have four same modular divisions, which are planning and control module, sensory module, actuation module and communication module.



**Fig. 2.** (Left) Software architecture; (Right) Flowchart of planning and control

The planning and control module is the main control system which plans the walking gait and converts it into joint commands. In every walking cycle, the planning is conducted two times during the dual single support phases and two for dual double support phases. The gait is planned on-line according to the feedback from sensors and modified by stability controller and landing controller. Then, the amended gait will convert into real-time joint control commands through inverse kinematics. The flowchart is shown in Fig. 2.

For sensory module, it manages all the sensors, process the signals and sends their data to the walking and control module. While for the communication module, it is the collection of the inter-process communication and the CAN bus communication. It monitors the communication status and maintains the communication infrastructure. The actuation module handles the local joint controllers. It obtains commands from the planning and control module and operates the CAN bus to change the motion status of joint controllers.



### 2.3 Landing Controller

The landing controller plays an important role in providing walking stability under uneven ground condition, during foot landing in phase 2 and 4 of walking cycle. The landing controller utilizes the readings from 6 axis force-torque sensor, in manipulation of the pitch and roll motions at ankle joint; to provide compliance motion to the unknown inclination and height of the ground. Termination of the foot landing occurs when the foot structure reported a firm foot-floor contact [11].

## 3 Development of Foot Mechanism

This study adopted a low cost approach in developing foot mechanism that not only provide a mechanically stable configuration under uneven terrain condition, but also collect enough information for controller in foot landing decision. By studying at human feet and walking biomechanics, features for stable walking are extracted to be applied on this foot mechanism.

Before starting on the detail designs of the foot mechanism, the basic dimension and silhouette was first determined to serve as a reference and starting point of the design. In order to create a biped robot that is human-like, anthropometric data was referred and used in the design [12]. Taking the anthropometric data of Asian, Japanese men to be specific, the mean body height is 1688mm while foot length and breadth are 251mm and 104mm respectively. Since the height of LOCH robot is 1900mm, taking the foot size proportionately, the dimension of LOCH foot should be 282mm x 117mm. Conceptual design for the foot was started from this dimension, however, due to installation of mechanical components and safety consideration, the final dimension of foot developed is 300mm x 200mm.

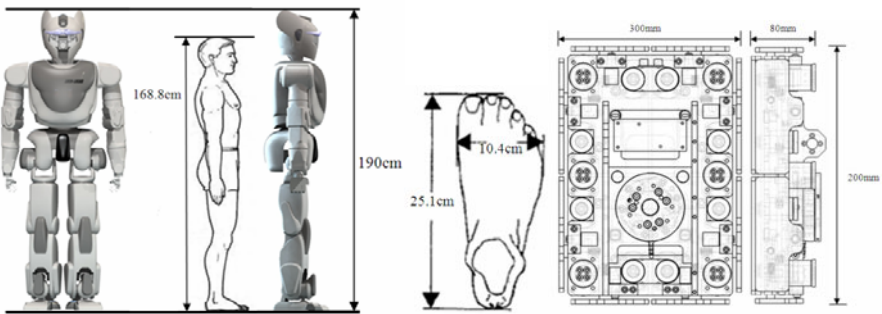


Fig. 3. Comparison of human anthropometric data with LOCH robot size

### 3.1 Design Specifications

Human rely on sensomotoric systems like vestibular organs, proprioceptive receptors (muscle spindles, stretch reflexes), exteroceptive tactile cues (pressure) and vision, to maintain dynamic balance during locomotion. These sensory feedbacks are rapidly processed by the central nervous system with high accuracy [13]. The mapping of

human sensory organs above to mechatronics hardware is as below: Vestibular organs balance human posture; provide sense of spatial orientation and information about linear and rotational movement. It can be represented by gyroscope. Proprioceptive receptors provide sense of the relative position of neighboring parts of body. It can be represented by strain gauge. Exteroceptive tactile cues provide sense of touch and pressure. It can be represented by touch or pressure sensor. Vision system extracts information from surrounding through eye. It can be represented by camera.

Taking away vision system and vestibular organs which are not available on foot, we can conclude the specifications as, (I) Exteroceptive to detect foot landing, (II) Exteroceptive to detect obstacles, (III) Proprioceptive to sense forces and torques acting on the foot. Besides that, the structure of human foot contributes to safe and comfortable locomotion, where arches of foot support the body weight and absorb impact, and the friction ridges at sole improves the roughness of skin and provides better grip onto the ground. Hence, two more specifications required are (IV) Shock absorption mechanism, (V) Sufficient grip at sole.

In order to accommodate for the workspace and kinematic design of the biped, the foot structure must also satisfy the following conditions: (VI) Maximum heel-contact and toe-off motion of  $30^\circ$ , (VII) Maximum walking speed of  $0.3\text{ms}^{-1}$  and maximum step length  $0.5\text{m}$ , (VIII) Weight of foot should not be too heavy with maximum permissible of  $4\text{kg}$ , (IX) Able to operate under light drizzle environment.

### 3.2 Selection Criteria

**Criteria for shock absorption.** Considering worst case falling scenario during walking, the required degree of shock absorption can be determined. The worst case scenario happens during single support phase, where the lifting foot reaches highest point and swinging with maximum speed. Hence, the maximum impact energy,

$$E_I = 0.5mv^2 + mgh . \quad (1)$$

Leg swinging phase for single support is normally taken to be 40% of the cycle time. Hence, for maximum walking speed,  $V_{\max}$  and maximum step length,  $L_{\max}$ , the maximum leg swinging speed,

$$V_S = 0.4L_{\max}/V_{\max} . \quad (2)$$

Thus, taking maximum value of  $4\text{kg}$  for the weight of the foot and  $5\text{cm}$  for the step height, the maximum impact energy is  $2.85\text{J}$ .

In order to prevent damages onto the leg especially the harmonic drives which coupled at the forefront of ankle joints, the shock absorption mechanism on foot must be able to absorb minimum of  $2.85\text{J}$ .

**Criteria for safe friction.** (walking under wet condition) Walking is considered as safe when the coefficient of friction,  $\mu$  is bigger than the ratio of horizontal force,  $F_H$  to the vertical force,  $F_V$  applied to the ground by biped feet, without considering the transverse force acting from the side.

$$\mu > F_H/F_V . \quad (3)$$

Frictional force,  $F_\mu$  is proportional to the normal force,  $F_N$ :

$$F_\mu = \mu F_N. \quad (4)$$

Then two equations combined:

$$F_\mu / F_N = F_H / F_V. \quad (5)$$

From here, we can evaluate the potential slippery condition and check the boundaries of acceptable parameters. Through this concept, we can further deduce the effect of coefficient of friction,  $\mu$  on determination of step length,  $s$  and leg length,  $l$  under equilibrium of forces condition. From geometrical analysis, we can get leg angle,

$$\theta = \sin^{-1}(s/2l). \quad (6)$$

and the relationship between leg angle and coefficient of friction is

$$\tan\theta = F_H / F_V = F_\mu / F_N = \mu. \quad (7)$$

Hence, the relationship between coefficient of friction with step length and leg length is

$$\mu = \tan(\sin^{-1}(s/2l)). \quad (8)$$

For human case, the required friction for normal walking is found to be lying between 0.15 to 0.30, with average on 0.17 for no skidding, 0.13 for slip-stick motion and 0.07 for falling result [13].

**Criteria for body structure.** During the heel-contact or toe-off phase, there will be huge shearing force acting horizontally in the foot structure. With LOCH's weight of 95kg and maximum angle of inclination  $30^\circ$ , the shearing force,  $F_{\text{shear}}$  can be determined from geometrical relationship, which has maximum value of 1863.9N.

$$F_{\text{shear}} = W / \sin 30^\circ. \quad (9)$$

## 4 Hardware Design

The hardware design of the LOCH feet is divided into three layers. First layer is connected to the bottom part of six axis force-torque sensor and carry the six main pillars of connections with middle plate. The middle layer is the backbone of the foot which holds mechanical switches while bottom layer consists of four separate plates which can move in vertical direction separately to detect firm landing of robotic foot.

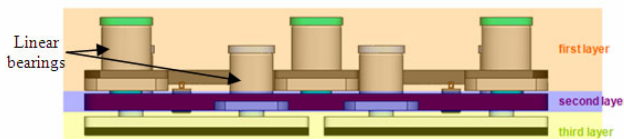


Fig. 4. Side view of foot design showing three layers of vertical moving plates

#### 4.1 Shock Absorption Elements

The foot structure is separated into three layers, in order for vertical movement between layers to achieve shock absorption and ground detection purposes. Between top and middle layers, eight coil springs of spring constant 78.1N/mm and maximum deflection of 3.2mm are chosen to provide potential energy of 3.2J upon full compression. The safety factor taken (which is around 1.2) is relatively small; in order to prevent the high stiffness of coil spring over-damps the shock absorption system. With outer forces exerted on feet absorbed by the coil springs, the damage done on biped structure especially harmonic drive will be reduced to minimum level.

Coil springs are placed in holes of diameter 1.5mm bigger than their original diameter to avoid friction between springs and hole wall which will weaken thus break the springs after repetitive movement. To position the springs in place, pre-load force is applied by top and middle layer of the foot structure by limiting top structure moving range.

However during standing position, due to high spring force, the weight of robot will not be able to compress the coil springs fully and it may cause the robot to wobble when acted by a force from the side. The instability is thus overcome by the usage of linear bearings which restrict the sliding movement of the upper plate and lower plate.

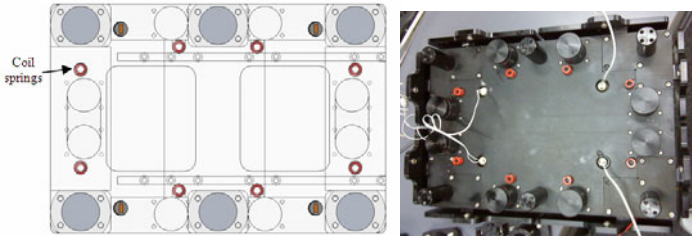


Fig. 5. Spring position at middle layer of foot design

#### 4.2 Friction Design

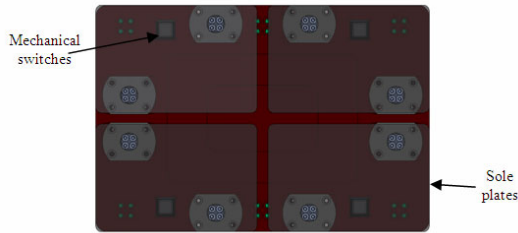
Referring to the criteria for safe friction above, the design of the leg is checked for the required maximum step length. While LOCH has leg length of 1.02m, the specified maximum step length of 0.5m gives a coefficient of friction of 0.25, which is safe for non-skidding movement.

In order to further improve the grip force, rubber pads are attached to the sole at bottom layer. According to ANOVA testing of wet elastomer, rubber possesses coefficient of friction of more than 0.4 when tested sliding on wet floor [14]. Even when lubricated with oil, it still shows a value of around 0.2. Among all type of rubbers, nitrile butadiene rubber (NBR) is chosen due to its good performance on abrasion resistance, water and oil resistance, chemical resistance and also anti-aging properties; on top of its low cost compared to fluorine rubber (FPM). Besides that, the advantages of adding rubber pads include small conformance to the irregularities of the uneven ground due to compressibility of rubber and enhancement of shock absorption capability.

### 4.3 Landing Detection

In order to provide inputs for the control algorithm of walking on uneven ground, mechanical switches are installed onto each foot to give the contact information of robotic feet with ground. The inputs required here is only a discrete on-off signal where momentary pushbuttons are used, which connect the circuit only when it is pressed. They are installed between middle and bottom layer. The number of switches to be mounted onto the bottom of feet can be maximized to cover entire area of feet. However, for simplicity and stability in mechanical assembly, the landing of foot can be examined from four points preferably four corners of the foot. Too complicated design may cause the movement between bottom layer and middle layer jammed and thus defeat the purpose of the design.

With four mechanical switches, each foot is separated into four portions: left front, right front, left rear and right rear. Four plates are mounted at each portion to aid four respective switches as shown in figure. The installation of plates can extend the surface area covered by one switch to near one quarter of the area of foot, which means that any size and shape of irregularities on the ground first touch the plate will activate the mechanical switch and indicates an “on”. In order to determine a firm landing on the ground, three activations of mechanical switches are enough to ensure full contact of feet with ground. The stability controller will then monitor the zero moment point to ensure it lies within pre-defined stable region under the foot.



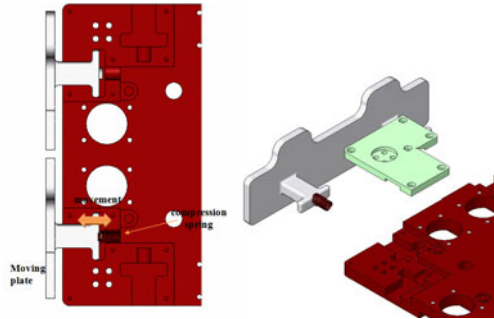
**Fig. 6.** Mechanical switches at the middle layer with four moving sole plates forming bottom layer

The usage of four plates not only maximizes the area covered by each switch, but also protects the switches from direct damage and abrasion. Protruding the mechanical switches out from the bottom of feet functioning as four supporting points for the feet will impose high impact when landing of foot, in which may damage and spoil the pushbuttons. Hence, by exposing only the moving button parts and keeping the body of switches inside the feet, the impact of landing will only be exerted directly onto the foot structure without passing through the switches. Furthermore, with the rubber pad attached onto four plates, slippage of the foot can be prevented.

### 4.3 Obstacle Detection

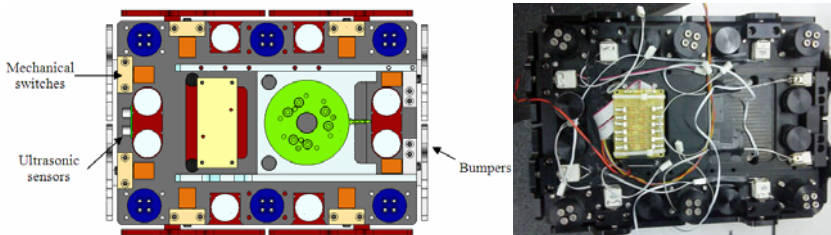
During the planning of walking motion, LOCH robot will inevitably come across obstacles. Instead of using visual guidance to identify obstacles in front, a more direct way to handle the situation would be through mechanical and electronics means.

There are eight mechanical switches installed onto LOCH foot, with two switches mounted at each side to detect the collisions of foot with obstacles. Similarly, in order to avoid direct collision of switches onto obstacles which may induce high impact force and damage the switches, metal plate are used as bumper at the frontline of collision. Compression springs are mounted at the back of the plates to recover their position after collision while the metal plates are positioned in sliding grooves of middle layer. The movement range of metal plates is defined by their T-shape design at the back.



**Fig. 7.** Bumper mechanism for side obstacle detection

Ultrasonic sensor is installed at the robotic foot to gauge the distance of obstacles in front. The ultrasonic sensors used are SRF-10 from Devantech which have compact size and wide detection range. During the landing phase of the foot, ultrasonic sensor will be activated to check for obstacles in front. If the distance is less than the walking step length, the walking speed and leg swinging speed will be reduced while ultrasonic sensor continue to monitor the distance. When the foot has collided the obstacle from the feedback of mechanical switch, obstacle crossing algorithm will be activated to walk across or walk pass it.



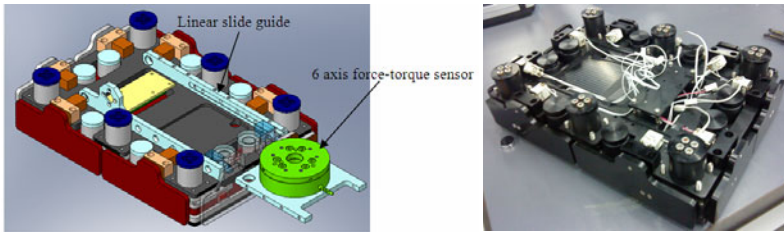
**Fig. 8.** Top view showing mechanical switches with bumper activation mechanism

#### 4.4 Mechanical Components

For the vertical motion, instead of direct vertical sliding like the plate for bumper, linear bearings are used as the main pillars to support the vertical movements between different layers. Since the shearing force during heel-contact or toe-off phase have maximum value of 1863.9N, six main linear bearings of diameter 20mm with

dynamic load rating of 882N each are installed between top and middle layers; while eight linear bearings of diameter 16mm with dynamic load rating of 775N are mounted onto middle and bottom layers. Higher safety factor of around 3.5 are adopted here because linear bearings are critical links between different layers, and the actual shear force may be higher when suffered from collision impact.

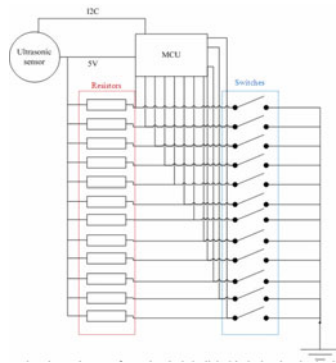
Besides that, to ease assembly, regular inspection and maintenance, the foot structure is designed as an individual module that can be easily separated from ankle joint. The design utilizes a plug and play method where the foot module can be easily slid into the position and locked using screws. Since the screws are applied horizontally, this method shifts the major robot weight onto the linear guide instead of exerting directly onto screw, when the screws are locked vertically. This aid in the stability of the joint and also increase the working life of fasteners.



**Fig. 9.** Foot design as a sub-module for easy assembly onto ankle joint using linear sliding mechanism. The 6 axis force-torque sensor is directly fastened to ankle joint

### 4.5 Electronics Components

The electronic design on the foot is relatively simple with Avr microcontroller to gather input from twelve mechanical switches and gather data from ultrasonic sensor through I2C bus. The schematic diagram of the circuit is shown in Fig. 11.



**Fig. 10.** Schematic design of electronic circuit on LOCH feet

Since the requirement of operations includes light drizzle situation, the circuit is glued with heat glue and protected with tape for minimum water resistance capability.

## 5 Conclusion

In this paper, the design criteria and detail design of foot mechanism for LOCH robot has been presented. The first prototype confirms more stable walking gait with shock absorbing elements installed and basic compliance motion was achieved. The drawback of the design found is due to the NBR rubber pad used, where deformable rubber pad caused the robot to wobble left-right and front-back. Thinner rubber pad will be used with further experiments to check for the best configurations, and the result will be discussed in future publications.

## References

1. Suwanratchatamane, K., Matsumoto, M., Hashimoto, S.: A Simple Tactile Sensing Foot for Humanoid Robot and Active Ground Slope Recognition. In: IEEE International Conference on Mechatronics, April 14-17, pp. 1–6. IEEE Press, New York (2009)
2. Takashi, Y., Nishiwaki, K., Kagami, S., Mizoguchi, H., Inoue, H.: High Speed Pressure Sensor Grid for Humanoid Robot Foot. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3909–3914 (2005)
3. Yang, H., Shuai, M., Qiu, Z., Wei, H., Zheng, Q.: A Novel Design of Flexible Foot System for Humanoid Robot. In: IEEE Conference on Robotics, Automation and Mechatronics, pp. 824–828 (2008)
4. Minakata, H.: A Study of Flexible Shoe System for Biped Robot. In: The 8<sup>th</sup> IEEE International Workshop on Advanced Motion Control, pp. 387–392 (2004)
5. Hashimoto, K., Sugahara, Y., Ohta, A., Sunazuka, H., Tanaka, C., Kawase, M., Lim, H., Tanakashi, A.: Realization of Stable Biped Walking on Public Road with New Biped Foot System Adaptable to Uneven Terrain. In: The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechanics, pp. 226–231 (2006)
6. Hashimoto, K., Hosobata, T., Sugahara, Y., Mikuriya, Y., Sunazuka, H., Kawase, M., Lim, H., Tanakashi, A.: Development of Foot System of Biped Walking Robot Capable of Maintaining Four-point Contact. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1361–1366 (2005)
7. Li, J., Huang, Q., Zhang, W., Yu, Z., Li, K.: Flexible Foot Design for a Humanoid Robot. In: IEEE International Conference on Automation and Logistics, pp. 1411–1419 (2008)
8. Yamaguchi, J., Takanishi, A., Kato, I.: Development of a Biped Walking Robot Adapting to a Horizontally Uneven Surface. In: Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems 1994. Advanced Robotic Systems and the Real World, vol. 2, pp. 1156–1163 (1994)
9. Shimizu, H., Wakazuki, Y., Pan, Y., Furuta, K.: Biped Walking Robot Using a Stick on Uneven Ground. In: Annual Conference SICE, pp. 83–88 (2007)
10. Xian, L.B., Xie, M.: Design of Human-like Leg-foot for Human-assisted Biped Walking. In: Xiong, C.-H., Liu, H., Huang, Y., Xiong, Y.L. (eds.) ICIRA 2008. LNCS (LNAI), vol. 5314, pp. 520–527. Springer, Heidelberg (2008)
11. Zhang, G., Xie, M., Yin, H., Wang, L., Yang, H.: Planning and Control of Biped Walking along Curved Paths on Unknown and Uneven Terrain. In: Xie, M., Xiong, Y., Xiong, C., Liu, H., Hu, Z. (eds.) ICIRA 2009. LNCS, vol. 5928, pp. 1032–1043. Springer, Heidelberg (2009)
12. Marras, W.S., Karwowski, W.: Fundamentals and Assessment Tools for Occupational Ergonomics. CRC Press, Boca Raton (2006)
13. Kumar, S.: Biomechanics in Ergonomics. Taylor & Francis, Philadelphia (1999)
14. Smith, R.H.: Analyzing Friction in the Design of Rubber Products. CRC Press, Boca Raton (2008)



# 4WD Skid-Steer Trajectory Control of a Rover with Spring-Based Suspension Analysis Direct and Inverse Kinematic Parameters Solution

Edgar Martínez-García and Rafael Torres-Córdoba

Robotics Laboratory  
Institute of Engineering and Technology  
Universidad Autónoma de Ciudad Juárez, México

**Abstract.** This manuscript provides a solution to the problem of four-wheel drive (4WD) kinematics and dynamics for trajectory control of an in-wheel motors rover. The rover is a platform built up in our laboratory, featured by its damper devices. The rover wheels' contact point have dynamic positions as effects of spring-based suspension devices damper. The tracking control is a four-wheel drive skid-steering (4WDSS) system, and we propose a motion control fundamentally defining a dynamic turning  $z$ -axis, which moves within the area of the four wheels' contact point. We provide a general solution for this mechanical design since the wheels' contact point displacement directly impacts the rover angular velocity. Furthermore, we introduce a model for inertial localization based on an arrangement of two accelerometers to define the rover position within a global inertial frame.

## 1 Introduction

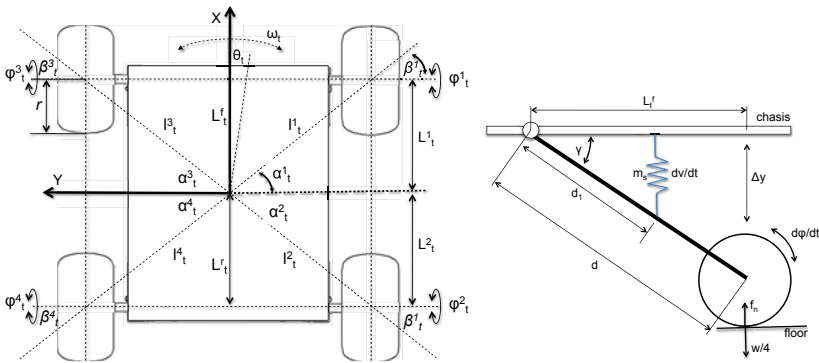
This manuscript provides a formulation in continuous time to calculate inverse and direct kinematics parameters for trajectory control of a four-wheel drive (4WD) mobile robot with in-wheel motors and independent suspensions (fig. 1) to exploit the skid-steering ability. One the problems stated here is defined as how to describe trajectory control with asynchronous wheels drive based on skid-steering (4WDSS) [7][8][5], since robot angular acceleration  $\omega_t$  works different with respect to (w.r.t.) other traditional synchronous drive control systems. The angular acceleration will depend on the wheels position overtime as the damper device cause effects in  $L_1$  and  $L_2$  (fig. 1). The authors present three different general solutions for the angular speed: a) under no damper constraints in totally plain terrains; b) under spring-mass effects in plain floors; and c) under uneven terrains where the normal force is considered. Furthermore, the models presented include dead-reckoning analysis with kinematic and dynamic parameters solutions, which are used to characterize real trajectory control. Many trajectory control research reported in the related work present approaches to control yaw using different sensing modalities with fundamentals in the rover kinematics of non-holonomic systems [7], [1], [4], skid control properties [8] [10], and direct

control over the effectors [3]. Other works, include to this problem the ability to control position and angle steer to achieve tracking control [2], [6], or preserving stability [5], and in some cases obstacle avoidance is also complemented [1]. Our approach combines simple low cost devices to accomplish trajectory tracking and inertial localization for all terrains. The section 2 describes the mechanical design and kinematic parameters of the 4WD rover. Section 3 details the rover angular velocity model under different kinematic conditions. The section 4 describes in detail how the suspension influences the wheels' contact point locations, and how such model is integrated with the angular velocity model. The section 5 describes the trajectory control model and the dynamic rotation axis w.r.t. the instantaneous center of rotation. Section 6 provides the parameter solution for the 4WDSS inverse kinematics. Section 7 is devoted to the inertial unit model for inertial localization, describing the results on experimental implementations. Finally, section 8 provides the paper conclusions.

## 2 Robotic Platform Kinematic Configuration

According to fig.1 a four-wheel drive skid-steer (4WDSS) robotic platform with independent wheels control speeds  $\dot{\varphi}_t^i$  is analysed. The distance  $l_t^i$  between wheels and the geometric center is variable on time as the robot moves with angles  $\alpha_t^i$ . The real distance between contact points on one robot's side is defined by  $L_t^i + L_t^{i+1}$ . The rover's angular velocity and bearing direction are defined by  $\omega_t$  and  $\theta_t$  respectively.

The angle  $\beta_t$  is non-stationary and represents a kinematic restriction featured by the angle between the rotational wheel axis, and the line which cuts the rover center of mass and the wheel contact point [9]. The rover inertial system is defined locally by the  $X$  along its longitudinal axis, and  $Y$  in its transversal axis. There exist a mechanical arm of length  $d$  for each wheel, with two joints. On one extreme, the suspension is fixed around the middle rover side with a rotative



**Fig. 1.** Left: Top view of the rover-like mobile robot with its kinematic configuration. Right: Kinematic configuration of the spring-based suspension of each wheel.

mechanical device that allows free angular motion. The another joint is a spring-based damper device located around the middle of the arm tied vertically to the chassis plate (as depicted in fig.2-right). The angle  $\gamma_t^{f,b}$  represents the angular position of each suspension arm, measured with a linear transducer implemented in the rotative arm device.

### 3 Angular Speed Models

Since the rover has independent wheels speed control, its instantaneous velocity model is defined by the averaged wheels’s speeds as given by the equation (1),

$$v_t = \frac{r}{4} \sum_{i=1}^4 \dot{\varphi}_t^i \tag{1}$$

Where  $r$  represents the nominal wheels radius (assuming in this manuscript all wheels have the same  $r$  magnitude), and  $\dot{\varphi}_t^i$  is each of the rotational wheels’ speed. Furthermore, the rover’s differential linear velocity  $\hat{v}_t$  is formulated as in the expression (2),

$$\hat{v}_t = r(\dot{\varphi}_t^1 + \dot{\varphi}_t^2 - \dot{\varphi}_t^3 - \dot{\varphi}_t^4) \tag{2}$$

where left-sided wheel rotational speeds are denoted to have minus sign w.r.t. the right-sided wheels. Therefore, the general model for the robot angular velocity is defined respect to its center of mass, ideally in its geometrical center by

$$\omega_t = \frac{\hat{v}_t \cos(\theta_t)}{L_t^f} = \frac{r \cos(\theta_t)}{L_t^f} (\dot{\varphi}_t^1 + \dot{\varphi}_t^2 - \dot{\varphi}_t^3 - \dot{\varphi}_t^4) \tag{3}$$

$\theta_t$  is the bearing robot direction at continuous time  $t$ , and the reference point of rotation is taken from the middle front side of the robot namely  $L_t^f$ . Also, representing the contact point projection over the front side longitudinal vehicle axis. Now, we can easily define the control vector by

$$\mathbf{u}_t = \begin{pmatrix} v_t \\ \omega_t \end{pmatrix} = \begin{pmatrix} \frac{r}{4} & \frac{r}{4} & -\frac{r}{4} & -\frac{r}{4} \\ \frac{r \cos(\theta_t)}{L_t^f} & \frac{r \cos(\theta_t)}{L_t^f} & -\frac{r \cos(\theta_t)}{L_t^f} & -\frac{r \cos(\theta_t)}{L_t^f} \end{pmatrix} \begin{pmatrix} \dot{\varphi}_t^1 \\ \dot{\varphi}_t^2 \\ \dot{\varphi}_t^3 \\ \dot{\varphi}_t^4 \end{pmatrix} \tag{4}$$

where  $\mathbf{u}_t = (v, \omega)^T$  is control vector by which the robot is moved toward a desired local goal destination in terms of direct kinematics [13,11]. While next expression defines velocity state control vector but in terms of inverse kinematics,

$$\dot{\xi}_t = \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} \cos \theta_t & 0 \\ \sin \theta_t & 0 \\ 0 & 1 \end{pmatrix} \mathbf{u}_t \tag{5}$$

### 3.1 Case for No Spring-Mass Model

The following is a set of equations for modelling  $\omega_t$  with kinematics motion only. For this case the suspension has no energy release because the rover moves around a totally man-made flat floor. Firstly, let us define a general solution for  $\alpha_t^i$  by (see fig. 10),

$$\sin(\alpha_t^i) = \frac{L_t^f}{l_t^i}; \quad L_t^f = d \cos(\gamma_t^i) \tag{6}$$

It follows a formulation of linear equations to project  $L_t^i$ . By knowing a model for  $L_t^f$ , wheels position can be inferred to calculate the robot angular velocity.  $L_t^f$  is substituted to define a first equation,

$$l_t^i \sin(\alpha_t^i) = d \cos(\gamma_t^i); \quad l_t^i \sin(\alpha_t^i) - d \cos(\gamma_t^i) = 0 \tag{7}$$

likewise, second equation is defined

$$\frac{W}{2} \tan(\alpha_t^i) = d \cos(\gamma_t^i); \quad \frac{W}{2} \tan(\alpha_t^i) - d \cos(\gamma_t^i) = 0 \tag{8}$$

and the third equation is also defined by

$$\frac{W/2}{\cos(\alpha_t^i)} = d \cos(\gamma_t^i); \quad \frac{W}{2 \cos(\alpha_t^i)} - d \cos(\gamma_t^i) = 0 \tag{9}$$

For (9) we define an equation in terms of  $\alpha_t^i$ ,

$$\alpha_t^i = \frac{W}{2 \cos(\gamma_t^i)} \tag{10}$$

Similarly, for calculating  $l_t^i$  using (7) and (8), and algebraically arranging,

$$l_t^i = \sin(\alpha_t^i) = \frac{W}{2} \tan(\alpha_t^i); \quad l_t^i = \frac{W \tan(\alpha_t^i)}{2 \sin(\alpha_t^i)} \tag{11}$$

with  $l_t^i$  dropped off, let us substitute functional form of  $\alpha_t^i$

$$l_t^i = \frac{W \tan(\arccos(\frac{W}{2d \cos(\alpha_t^i)}))}{2 \sin(\arccos(\frac{W}{2d \cos(\alpha_t^i)})} \tag{12}$$

Thus, the following model is a limit case where no spring-damper is considered, including a definition for  $\kappa$ ,

$$\omega_t = \frac{2r \cos(\theta_t)}{\kappa} \sum_{i=1}^4 \dot{\varphi}_t^i; \quad \kappa = \frac{W \tan(\alpha_t^i)}{2 \sin(\alpha_t^i)} \alpha_t^i \tag{13}$$

## 4 Suspension Motion Analysis

The general spring-mass formulation is defined by (14) according to the geometric suspension design setup as a vertically aligned spring-based device. The model regards two constants in the equation, the spring restitution coefficient  $\kappa_1$ , and the damper coefficient  $\kappa_2$ .

$$F_t = -\kappa_1 \Delta_y - \kappa_2 \dot{y}^n + \frac{w}{4} \quad (14)$$

Where,  $F_t$  is the force required to exert spring linear motion at time  $t$ , its motion yields a displacement vertically namely  $\Delta_y$  with a spring friction constant  $\kappa_1$ . The spring-mass damper term is  $\kappa_2 \dot{y}^n$ , with  $n = 1$  which is a linearisation parameter. This term is associated to the spring instant linear velocity. The vehicle weight  $w$  is divided by four-wheel contact points, so that the right-sided term of the equation (14) is denoted by  $w/4$ . The spring device displacement is defined by (15)

$$\Delta_y = d_1 \cos(\gamma_t^i) \quad (15)$$

Thus, in terms of spring-mass and acceleration,

$$m_s \ddot{y}_t = -\kappa_1 d_1 \cos(\gamma_t^i) - \kappa_2 \dot{y}^n + \frac{w}{4} \quad (16)$$

Where  $m_s$  is the spring mass, and  $a_t$  is the acceleration motion yielded by the spring damping effects. Algebraically arranging, let us substitute (16) into the equation (6) in terms of  $\cos(\gamma_t^i)$ , thus

$$\cos(\gamma_t^i) = \frac{m_s \ddot{y}_t + \kappa_2 \dot{y} + \frac{w}{4}}{-\kappa_1 d_1} \quad (17)$$

### 4.1 Condition for Spring Damper

The condition for motion over totally even terrains or structured floors, the normal force is declared as  $f_n = 0$  because of consideration of no spring motion is regarded. Therefore, substituting in terms of  $L_t$  with  $f_n = 0$ ,

$$L_t^f = \frac{d}{-\kappa_1 d_1} (m_s \ddot{y} - \kappa_2 \dot{y} - \frac{w}{4}) \quad (18)$$

Hence, the vehicle angular rotation speed constrained by its dynamic wheels contact point is presented by next equation,

$$\omega_t = \frac{-\kappa_1 d_1 r \cos(\theta_t)}{d(m_s \ddot{y} - \kappa_2 \dot{y} - \frac{w}{4})} (\dot{\varphi}_t^1 + \dot{\varphi}_t^2 - \dot{\varphi}_t^3 - \dot{\varphi}_t^4) \quad (19)$$

## 5 Trajectory Control Model

For trajectory control we combine the instantaneous center of rotation (ICR), from which there exist a radius  $R$  to the robot rotation point coordinate  $(x_R, y_R)^T$ . Let us define a general assumption,

$$R = \frac{v}{\omega} = \frac{r \sum_i \dot{\varphi}_i^t}{4\omega_t} \tag{20}$$

The robot angular velocity  $\omega_t$  is now a known model that exert motion control effects over the robotic platform and can be defined for different cases:

1. Rigid or no suspension (fixed 4WD)

$$\omega_t = \frac{(\dot{\varphi}_t^1 + \dot{\varphi}_t^2 - \dot{\varphi}_t^3 - \dot{\varphi}_t^4) \cos(\theta_t)}{L_t^f}$$

2. Suspension with no damper effects (totally even floor and smooth turns)

$$\omega_t = \frac{4 \sin(\alpha_t)(\dot{\varphi}_t^1 + \dot{\varphi}_t^2 - \dot{\varphi}_t^3 - \dot{\varphi}_t^4) \cos(\theta_t)}{W \tan(\alpha_t^i) \alpha_t^i}$$

3. Suspension with spring-mass effects

$$\omega_t = \frac{-\kappa_1 d_1 r (\dot{\varphi}_t^1 + \dot{\varphi}_t^2 - \dot{\varphi}_t^3 - \dot{\varphi}_t^4) \cos(\theta)}{d(m_s \ddot{y} - \kappa_2 \dot{y} - w/4)}$$

However, the third case could be chosen as the most general case substituting the first two cases (if required). Our proposed strategy establishes that the robot trajectory is tracked by the robot placing over the path a robot rotation axis which is located at  $(x_R, y_R)^T$ . The rotation axis is dynamically positioned within the area of the four wheel contact points, depending on the inertial robot behaviour. Thus, this dynamic effect is fundamental in the 4WDSS nature of the robot (see also [5]). The rotation axis is then formulated within the robot's inertial fixed frame for  $XY$  axis,

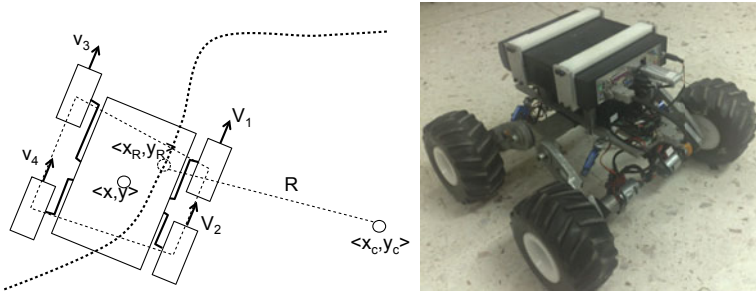
$$x_R = \frac{W}{v_{max}}(\dot{\varphi}_t^1 - \dot{\varphi}_t^2 - \dot{\varphi}_t^3 + \dot{\varphi}_t^4); \quad y_R = \frac{v_t L}{v_{max}}(-\dot{\varphi}_t^1 + \dot{\varphi}_t^2 - \dot{\varphi}_t^3 + \dot{\varphi}_t^4) \tag{21}$$

The rotation axis coordinates will move nearly around the robot centroid according to its motion direction. Thus, the ICR at coordinates  $\langle x_c, y_c \rangle$  is mathematically defined by,

$$\begin{pmatrix} x_c \\ y_c \end{pmatrix} = \begin{pmatrix} (x_t + x_R) - R \sin(\theta_t + \omega_t \Delta_t) \\ (y_t + y_R) + R \cos(\theta_t + \omega_t \Delta_t) \end{pmatrix} \tag{22}$$

By deducing an ideal robot position recursively w.r.t. the ICR model [13],

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} x_c - x_R - R \sin(\theta_t + \omega_t \Delta_t) \\ y_c + y_R + R \cos(\theta_t + \omega_t \Delta_t) \end{pmatrix} \tag{23}$$



**Fig. 2.** Left: 4WD skid-steer trajectory control and turning axis; Right: A photo of the rover-like mobile robot

## 6 Inverse Kinematic Parameters Solution

By means of inverse kinematics analysis wheels control parameters are solved, and trajectory control and stability are better defined [6]. The wheels kinematic parameters are featuring low level effector commands [10], hence robot location and robot velocities can be established for navigation control. Next is a set of linear equations defined in terms of direct kinematics.

$$\begin{aligned}
 \dot{\varphi}_t^1 + \dot{\varphi}_t^2 + \dot{\varphi}_t^3 + \dot{\varphi}_t^4 &= \frac{4v_t}{r} \\
 \dot{\varphi}_t^1 + \dot{\varphi}_t^2 - \dot{\varphi}_t^3 - \dot{\varphi}_t^4 &= \omega_t \frac{d(m_s \ddot{y} - \kappa_2 \dot{y} - w/4)}{-\kappa_1 d_1 r \cos(\theta_t)} \\
 \dot{\varphi}_t^1 - \dot{\varphi}_t^2 - \dot{\varphi}_t^3 + \dot{\varphi}_t^4 &= \frac{x_R v^{max}}{W} \\
 -\dot{\varphi}_t^1 + \dot{\varphi}_t^2 - \dot{\varphi}_t^3 + \dot{\varphi}_t^4 &= \frac{y_R v^{max}}{L}
 \end{aligned} \tag{24}$$

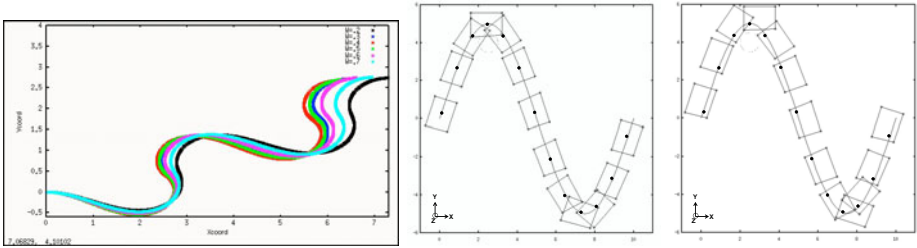
By solving the set of linear equations by Gauss-Jordan (or any other factorization method), we obtain the wheels velocity vector control  $\Omega_t = (\dot{\varphi}_t^1, \dot{\varphi}_t^2, \dot{\varphi}_t^3, \dot{\varphi}_t^4)^T$  which is the general solution for inverse kinematics,

$$\begin{aligned}
 \dot{\varphi}_t^1 &= \frac{v_t \lambda_1 + \omega_t \lambda_2 + x_R \lambda_3 - y_r \lambda_4}{4\lambda_1 \lambda_2 \lambda_3 \lambda_4} \\
 \dot{\varphi}_t^2 &= \frac{v_t \lambda_1 + \omega_t \lambda_2 - x_R \lambda_3 + y_r \lambda_4}{4\lambda_1 \lambda_2 \lambda_3 \lambda_4} \\
 \dot{\varphi}_t^3 &= \frac{-v_t \lambda_1 - \omega_t \lambda_2 - x_R \lambda_3 - y_r \lambda_4}{4\lambda_1 \lambda_2 \lambda_3 \lambda_4} \\
 \dot{\varphi}_t^4 &= \frac{v_t \lambda_1 - \omega_t \lambda_2 + x_R \lambda_3 + y_r \lambda_4}{4\lambda_1 \lambda_2 \lambda_3 \lambda_4}
 \end{aligned} \tag{25}$$

And defining the factors  $\lambda_i$  the next equation is described as,

$$\lambda_1 = \frac{1}{r}; \quad \lambda_2 = \frac{-\kappa_1 d_1 r \cos(\theta_1)}{d(m_s \ddot{y} - \kappa_2 \dot{y} - w/4)}; \quad \lambda_3 = \frac{v_{max}}{W}; \quad \lambda_4 = \frac{v_{max}}{L}$$

The purpose of this formulation is to calculate the wheels angular speed vector  $\Omega_t = (\dot{\varphi}_t^1, \dot{\varphi}_t^2, \dot{\varphi}_t^3, \dot{\varphi}_t^4)^T$ , which is essential to solve for the control vector  $\mathbf{u}_t$  in order to reach skid-steering trajectory control. Thus, by solving the linear equation system of (24) results are depicted in fig.4 that shows some simulated skid-steer based control trajectories (and empirically demonstrated in laboratory) similarly reported in [5].



**Fig. 3.** Left: Different trajectories yielded by changing robot width  $W$ ; Middle and Right: Robot trajectory controlled by parametrizing  $\Omega_t$  to shift the Z-turning axis  $\langle x_R, y_R \rangle$  denoted by the bold points

## 7 Inertial Localization

The main role of robot position regards as a parameter for the instantaneous center of rotation model in the trajectory control. A home made inertial module compounded of two accelerometers to measure displacement in a global framework was developed. Deploying inertial sensors for 4WDSS trajectory control works on a dead-reckoning modality, and wheel encoders [12] are not suitable for skid-steering techniques, where rough and uneven terrains are faced. The inertial module was built up with a couple of accelerometers, geometrically arranged at the front and rear side the robot, along the longitudinal robot axis, since only one accelerometer is able to measure respect to its fixed frame.

### 7.1 Global Frame Measuring Strategy

The angular acceleration is measured as a means to determine the total acceleration of a chassis point (see fig 4-left). The angular acceleration causes a tangential acceleration which is a component perpendicular to the position vector of such point, where the acceleration is being calculated with the rotation center of the body w.r.t. the origin. The orientation of the tangential acceleration bears toward a direction of the  $\alpha$  and perpendicular to a position  $\mathbf{r}$ , from the rotation center  $G$  of the robot, defined by next equation (a definition is described in [14]),



$$\mathbf{a}_{AG} = -\boldsymbol{\alpha} \times \mathbf{r}_A - \omega^2 \mathbf{r}_A; \quad \mathbf{a}_{BG} = -\boldsymbol{\alpha} \times \mathbf{r}_B - \omega^2 \mathbf{r}_B \quad (26)$$

Where  $\mathbf{a}_{AG}$  is the relative acceleration of point  $A$  (the front side) with respect to  $G$  (geometric center). The relation  $\omega^2 \mathbf{r}$  is also defined by  $\omega^2 \mathbf{r} = \omega \times \omega \times \mathbf{r}$ .  $\mathbf{a}_{BG}$  is the relative acceleration of point  $B$  (the rear side) with respect to  $G$ ,  $\mathbf{r}_A$  is the position vector of point  $A$  with respect to  $G$ ,  $\mathbf{r}_B$  is the position vector of point  $B$  with respect to  $G$ ,  $\boldsymbol{\alpha}$  is the angular acceleration vector of the robot, which is normal to the reference plane. The total acceleration of a point is equal to the acceleration of the rotation center  $G$ , plus the relative acceleration of the point with respect to  $G$ . Thus, for points  $A$  and  $B$  the following model can be stated,

$$\mathbf{a}_A = \mathbf{a}_G + \mathbf{a}_{AG}; \quad \mathbf{a}_B = \mathbf{a}_G + \mathbf{a}_{BG} \quad (27)$$

Where  $\mathbf{a}_G$  is the acceleration of the rotation center of the object. If total accelerations of  $A$  and  $B$  are subtracted, the acceleration of  $G$  is taken out, so this acceleration difference is independent of the chassis acceleration.

$$\mathbf{a}_B - \mathbf{a}_A = \mathbf{a}_{BG} - \mathbf{a}_{AG} \quad (28)$$

If equations (27) are substituted in (28) the result is as follows,

$$\mathbf{a}_B - \mathbf{a}_A = \mathbf{r}_B \times \boldsymbol{\alpha} + \omega^2 \mathbf{r}_B - (\mathbf{r}_A \times \boldsymbol{\alpha} + \omega^2 \mathbf{r}_A) \quad (29)$$

$$\mathbf{a}_B - \mathbf{a}_A = (\mathbf{r}_B - \mathbf{r}_A) \times \boldsymbol{\alpha} + \omega^2 (\mathbf{r}_B - \mathbf{r}_A) \quad (30)$$

After simplification it leads to:

$$\mathbf{a}_{AB} = \mathbf{r}_{AB} \times \boldsymbol{\alpha} + \omega^2 \mathbf{r}_{AB} \quad (31)$$

Being  $\mathbf{a}_{AB}$  the difference of accelerations of  $A$  and  $B$ . By definition,  $\mathbf{r}_{AB} \times \boldsymbol{\alpha}$  and  $\omega^2 \mathbf{r}_{AB}$  are orthogonal, so the following expression is valid.

$$\|\mathbf{a}_{AB}\| = \sqrt{(\|\mathbf{r}_{AB}\| \|\boldsymbol{\alpha}\|)^2 + (\|\mathbf{r}_{AB}\| \omega^2)^2} \quad (32)$$

If we solve for  $\|\boldsymbol{\alpha}\|$ , the following equation turns out,

$$\alpha = \|\boldsymbol{\alpha}\| = \sqrt{\left(\frac{\|\mathbf{a}_{AB}\|}{\|\mathbf{r}_{AB}\|}\right)^2 - \omega^4} \quad (33)$$

With this resulting equation is possible to calculate the angular acceleration of the moving object if the acceleration of two different geometric points are known. This formula includes the angular speed, which depends on the angular acceleration itself. As for implementation, such angular acceleration needs to be calculated iteratively using sensor readings. The following is a recursive

algorithm to calculate the angular acceleration, the angular speed, and the orientation angle of the robot,

$$\alpha_t = \sqrt{\left(\frac{\sqrt{(a_{xB} - a_{xA})^2 + (a_{yB} - a_{yA})^2}}{l}\right)^2} - \omega_{t-1}^4 \quad (34)$$

$$\omega_t = \omega_{t-1} + \int_t \alpha_t dt; \quad \theta_t = \theta_{t-1} + \int_t \omega_t dt \quad (35)$$

Where  $\alpha_t$  is the current angular acceleration of the robot.  $a_{xA}$ ,  $a_{xB}$ ,  $a_{yA}$  and  $a_{yB}$  are the acceleration values obtained from both two-axis accelerometers.  $l$  is the distance between the position of the accelerometers on the robot.  $\omega_t$  is the current angular speed of the robot.  $\Delta t$  is the time interval between each sensor reading. And  $\theta_t$  is the current angle of the robot.

## 7.2 Absolute Position

With the orientation angle being known, the acceleration and therefore the position of the system can be obtained according to an inertial reference frame, carried out by calculating the rotation of the accelerometers local frames, and integrating the result. The inertial coordinates of the front side accelerometer is given by

$$x_{AI} = \int_t \int_t (a_{xA} \cos\theta - a_{yA} \sin\theta) dt^2; \quad y_{AI} = \int_t \int_t (a_{xA} \sin\theta + a_{yA} \cos\theta) dt^2 \quad (36)$$

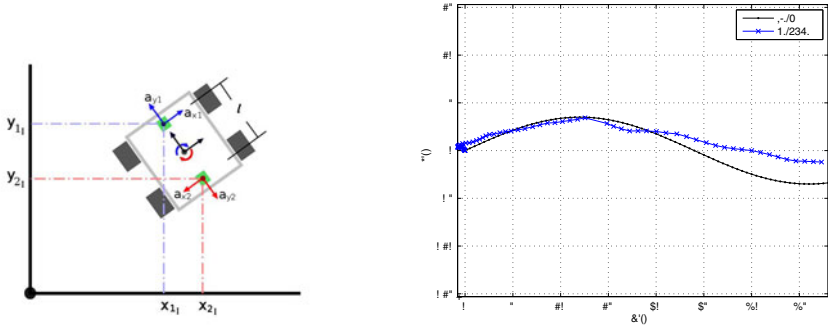
while the inertial coordinates of the rear side accelerometer are,

$$x_{BI} = \int_t \int_t (a_{xB} \cos\theta - a_{yB} \sin\theta) dt^2; \quad y_{BI} = \int_t \int_t (a_{xB} \sin\theta + a_{yB} \cos\theta) dt^2 \quad (37)$$

$x_I$  and  $y_I$  are the location of each accelerometer in an inertial frame, and where  $x_t$  and  $y_t$  are the current position of the robot,  $\theta_t$  is the current orientation angle,  $x_{t-1}$ ,  $y_{t-1}$ ,  $\theta_{t-1}$  are the last pose,  $v_t$  and  $\omega_t$  are the current linear and angular speeds, respectively, and  $\Delta t$  is the time interval between consecutive measures.

## 7.3 Experimental Analysis

The mobile robot used in this research work is a home made platform (see fig. 2 right). The rover can be configured in 4WDD (differential drive) or 4WDSS (asynchronous speeds) control modality. Experimentally, the robot was setup in tele-operation mode collect massive data, throughout an ideal curved trajectory (fig. 4 right), and computing the real position with the inertial localization formulation in real-time. The proposed two-accelerometer algorithm yielded position errors which increased as the robot navigates. Several main disturbance causes were experimentally identified that yielded error positions. Accelerometers sensitivity took an important roll despite of the home-made inertial unit



**Fig. 4.** Left: Array of two accelerometers instrumenting the 4WDSS rover. Right: Trajectory control in experimental results for  $\Omega_t$  and inertial localization under skid-steering motion control.

system capability to combine information. A couple of two low-cost accelerometers were the unite core. We found out that more accurate devices are required to reduce errors in the data fusion formulation. In earlier experiments, systematic causes were also detected to be a problem such as wheels mechanical shaking, and a suitable wheels sliding (lateral and longitudinal) formulation has not been integrated in this work yet. An optimal state estimation nonlinear filter for position measurement will be a matter of integration for the proposed model in the future. Likewise, a means based on GPS will be developed to reduce the accumulated error overtime within navigation segments where no speed changes are detected (although GPS was used, results are not depicted due to space in this manuscript).

## 8 Conclusions

A general solution to inverse and direct kinematics for trajectory control of a four-wheeled mobile robot was proposed. The mathematical formulation provided a model to explicitly control the robot’s angular velocity in terms of its dynamic contact points geometry and the wheels rotational rate, including in-wheel spring-based damper device. For the inertial localization, several main disturbance causes were experimentally identified that yielded error positions, such as accelerometers accuracy, the systematic causes such as wheels mechanical shaking. A suitable slipping formulation will be integrated, as well as position state estimation filtering. The paper provides a deterministic approach for three different damper conditions, establishing general solutions of the yaw speed. Finally, the presented framework has experimentally resulted feasible to be implemented in robotic platforms and trajectory control was stable enough for many navigational tasks.

## Acknowledgment

The authors would like to thank the partial support provided by PROMEP under project No. DGPD/SPROMEP/2-9/075.

## References

1. Ki, K., Wang, S., Di, H., Li, H.: Trajectory tracking control wheeled mobile robots based on integrated intelligent steering. In: IEEE ICEMI 2009, pp. 3-955-3-960 (2009)
2. Solea, R., Filipescu, A., Nunes, U.: Sliding-mode control for trajectory-tracking of a wheeled mobile robot in presence of uncertainties. In: Proc. of the 7th Asian Control Conference, Hong Kong, China, August 27-29, pp. 1701-1706 (2009)
3. Zhao, Y., Zhang, Y., Zhao, Y.: Stability control system for four-in-wheel-motor drive electric vehicle. In: IEEE 6th Intl. Conf. on Fuzzy Systems and Knowledge Discovery, pp. 171-175 (2009)
4. Silva-Ortigoza, R., Silva-Ortigoza, G., Hernandez-Guzman, V., Barrientos-Sotelo, R., Albarrán-Jiménez, Silva-García, V.: Trajectory tracking in a mobile robot without using velocity measurement for control of wheels. *IEEE Latin America Transaction* 6(7), 598-607 (2008)
5. Kozłowski, K., Pazderski, D.: Practical stabilization of a skid-steering mobile robot - A kinematic-based approach. In: IEEE Intl. Conf. on Mechatronics, pp. 519-524 (2006)
6. Caracciolo, L., De Luca, A., Iannitti, S.: Trajectory Tracking Control of a Four-Wheel Differentially Driven Mobile Robot. In: Proc. IEEE Intl. Conf. on Robotics and Automation, Detroit, Michigan, USA, pp. 2632-2638 (May 1999)
7. Mandow, A., Martínez, J.L., Morales, J., Blanco, J.L., García-Cerezo, A., González, J.: Experimental Kinematics for Wheeled Skid-Steer Mobile Robots. In: Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2007, San Diego, California, pp. 1222-1227 (2007)
8. Kozłowski, K.: Modelling and control of a 4-wheel skid-steering mobile robot. *Int. J. Appl. Math. Comput. Sci.* 14(4), 477-496 (2004)
9. Campion, G., Baston, G., Andrea-Novel, B.D.: Structural properties and classification of kinematic and dynamic models of wheeled mobile robots. *IEEE Transaction on Robotics and Automation* 12(1) (February 1996)
10. Wang, D., Low, C.B.: Modeling and analysis of skidding and slipping wheeled mobile robots: control design perspective. *IEEE Transactions on Robotics* 24(3), 676-687 (2008)
11. Macek, K., Thoma, K., Glatzel, R., Siewart, R.: Dynamics modeling and parameter identification for autonomous vehicle navigation. In: Proc. of the 2007 IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems, pp. 3321-3326 (2007)
12. Antonelli, G., Chiaverini, S.: Linear estimation of the odometric parameters for differential-drive mobile robots. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3287-3292 (2006)
13. Thrun, S., Burgard, W., Fox, D.: *Probabilistic Robotics*. MIT Press, Cambridge (2005)
14. Beer, F.P., Johnson, E.R., Eisenberg, E.R., Sarubbi, R.G.: *Vector mechanics for engineers dynamics*, ch. 15m, 5th edn., pp. 722-723. Mc Graw Hill, New York

# Topology and Analysis of Three-Phalanx COSA Finger Based on Linkages for Humanoid Robot Hands

Deyang Zhao and Wenzeng Zhang

Key Laboratory for Advanced Materials Processing Technology, Ministry of Education,  
Dept. of Mechanical Engineering, Tsinghua University, Beijing 100084, China  
wenzeng@tsinghua.edu.cn

**Abstract.** The design of a novel linkage three-phalanx finger was proposed in this paper. The finger is based on the concept of COSA (coupling and self-adaptation) and also two-phalanx COSA finger. As a combination of rigid coupled finger and self-adaptive under-actuated finger, the COSA finger has comprehensive grasping ability with human-like motion. Available topology designs were discussed in this paper and then mechanical design in details is given. Motion and grasping patterns of the finger were illustrated and compared with traditional coupled finger and under-actuated finger. Moreover, stability of grasp was simulated by static analysis. The three-phalanx finger with advantages of coupling and self-adaptation can be widely applied in hands for prosthetic limbs or humanoid robots.

**Keywords:** Humanoid robots, Under-actuated hands, Coupled finger, Self-adaptation, Linkage mechanism.

## 1 Introduction

In circumstance where grasp is in paramount concern, under-actuated hands were put forward as simplification of dexterous hands and application. Research in this field aims to create practical hands which are able to execute firm grasp more than precise manipulation. In an under-actuated hand, the number of actuators is lower than the number of degree of freedoms (DOFs)<sup>[1]</sup>. So that under-actuated hands have the advantages of simplicity and reliability. Fewer actuators and easier control strategy are utilized comparing to dexterous hands.

For under-actuated hands design, generally two approaches were proposed<sup>[2]</sup>. One is rigid coupling, which not accords with the definition of under-actuation though. Since the number of actuators is lower than the number of the active joints, compliant motion among joints is required. This solution is widespread in many designs such as, MA-NUS- Hand<sup>[3]</sup>, SDM Hand<sup>[4-5]</sup> and TBM Hand<sup>[6]</sup>. Due to the human-like features, the rigid coupled hands gains popularity in many humanoid robotic hands design. On the other hand, this type of finger lacks of adaptation to object geometry. A certain flexion pattern is imposed on rigid coupled fingers, predetermined by the mechanical design.

The second approach of under-actuated hands provides robot hands with self-adaptation or passive adaptation to object geometry. It is a crucial progress in this field. As a consequence, successful grasp likelihood is greatly improved in unknown

environment. Laval Hand<sup>[7-9]</sup>, TH-3R Hand<sup>[10]</sup> and GCUA Hand<sup>[11]</sup> all belong to this approach. This solution, however, has some disadvantages. The finger keeps straight and rotates as a rigid body when it is not in contact with the objects, which means it is not human-like. Moreover, if the object is not completely enclosed by the envelope of the fingers, the effect of the first contact might simply be to move the object away.

Apart from the two approaches, certain additions of the two were adopted in some robot hands. A three-phalanx finger mechanism is composed of seven links, one slider and two springs in total<sup>[12, 13]</sup>. In no contact motion, the middle and the distal phalanges rotate with respect to the proximal phalanx. The distal phalanx keeps parallel to the middle phalanx during this process. After the proximal phalanx is blocked by the cylinder, the middle and the distal phalanges contact the cylinder in sequence. Another three-phalanx finger was presented with six-bar linkage<sup>[14]</sup>. The proximal phalanx and the middle phalanx are jointed in self-adaptive way. The middle phalanx and the distal phalanx are rigid coupled. Both of the two fingers are simple addition of coupling and self-adaptation, since the two approaches are achieved separately on different parts of the finger.

The new concept, COSA (coupling and self-adaptation), combining the two approaches, is an innovated approach for robot hands design. This paper presents design of three-phalanx COSA finger in accordance linkage mechanism. The finger can be adopted in COSA hand with human-like motion and self-adaptive grasp.

## 2 Principle of Coupling and Self-Adaptation and Two-Phalanx COSA Finger

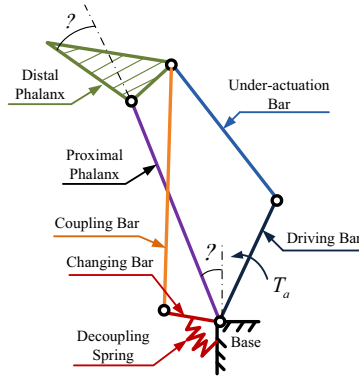
### 2.1 Principle of COSA

Rigid coupling and self-adaptive under-actuation are the two traditional actuation approaches, also the foundation of COSA. The rigid coupled finger moves through a specific path in space due to the relationship imposed between the angles of the phalanges, lacking of adaptation to object geometry as a result. In terms of self-adaptation, the unique motion of the finger leads to not only un-humanoid motion but also problems in grasping. Due to the straight configuration of the finger before contact, it is possible the enveloping of the object is not formed completely or firmly and then the grasp fails.

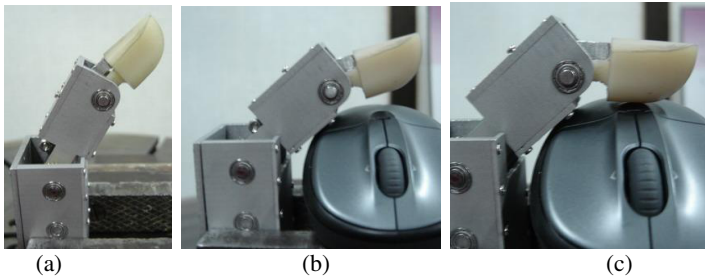
Aiming to solve the existing problem, COSA as a novel approach of under-actuation was put forward. It allows the robot finger to execute coupled and predetermined motion before any contact is made with the target, then self-adaptive motion after part of the finger is restrained by the target. A typical grasp by COSA finger is divided to two stages. In stage I that ends up with the contact, the finger flexes as a rigid coupled one. The ratio of rotating angle of one phalanx to another is constant, determined by the mechanical design. In stage II, the finger adapts to the object geometry “automatically”, actuating by the contact force.

### 2.2 Two-Phalanx COSA Finger Design

There are more than one types of mechanism to achieve COSA actuation. Available solutions include tendon cable, linkages, gear-racks and belt-pulley. Linkage



**Fig. 1.** Architecture of two-phalanx COSA finger i.e. COSA unit.  $T_a$ -the torque of motor.



**Fig. 2.** Two-phalanx COSA finger. (a) Coupled motion; (b) The end of grasping stage I; (c) The end of grasping stage II.

mechanism is selected in this research. Architecture of a two-phalanx COSA finger is shown in Fig. 1. The prototype of the finger is shown in Fig. 2. The grasp process is also demonstrated.

To achieve COSA, the most substantial issue here is to coordinate the motion exerted by the coupled mechanism and the under-actuated mechanism, respectively. The COSA finger should not only avoid conflict of the motions but also shift one motion to the other in proper timing. Decoupling elements are integrated to the finger including changing bar and decoupling spring in order to decouple the coupling mechanism and enable the under-actuation mechanism. The architecture equates to two four-bar linkage and a decoupling spring. The driving bar, the under-actuation bar, the distal phalanx and the proximal phalanx compose the self-adaptation four-bar linkage. The changing bar, the coupling bar, the distal phalanx and the proximal phalanx compose the coupling four-bar linkage. The decoupling spring, as decoupling element here, connects the changing bar to the base of the finger.

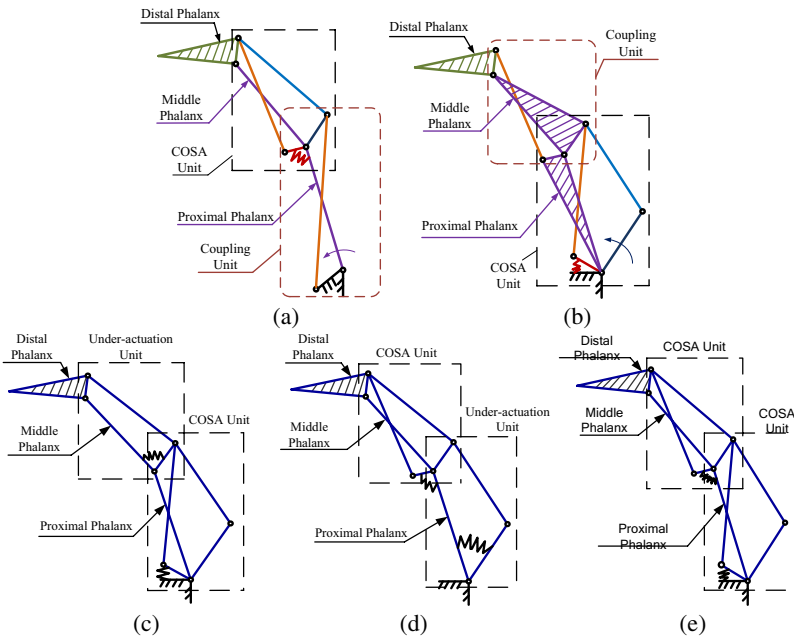
In stage I, the coupling bar and the changing bar act as fixed to the base during the finger is rotating. It is attributed to the constraint of the decoupling spring. The coupling

bar pushes the distal phalanx to rotate with respect to the proximal phalanx. In stage II, the under-actuation bar acts on the terminal of the coupling bar. The decoupling spring deforms while the coupling bar moves with the changing bar. The distal phalanx continues rotating until the object is enclosed, the grasp ended. During the motion, the changing bar and the decoupling spring are the key elements to separate the two stages. Thus, COSA actuation is achieved between two phalanges. This architecture is also called COSA unit, since it can be used to form  $n$ -phalanx COSA actuation.

### 3 Design of the Three-Phalanx COSA Finger

#### 3.1 Topology Designs

The three-phalanx COSA finger is designed based on the two-phalanx COSA finger i.e. the COSA unit in context of geometry. The COSA unit is adopted to achieve COSA actuation between two phalanges of three. It is not necessary that all the three phalanges adopt COSA. The coupling unit and the under-actuation unit are available for the finger to increase the flexibility of the design. Different combinations of these units create corresponding geometric finger designs with different feature. The finger generally includes two units. One unit links the distal phalanx and the middle phalanx and the other links the middle phalanx and the proximal phalanx. Available designs are shown in Fig. 3.



**Fig. 3.** Topology design of three-phalanx fingers. (a) COSA+Coupling; (b) Coupling+COSA; (c) Under-actuation+COSA; (d) Under-actuation+COSA; (e) COSA+COSA.



A connecting bar transmits the motion from the lower unit up to the upper unit. The length of the connecting bar is the key in the design since it is coupled with both two units. It influences the motion on each phalanx. The two units should be optimized so that both of them will work effectively.

In Fig. 3a, a COSA unit is used between the distal phalanx and the middle phalanx while the middle phalanx and the proximal phalanx are linked by the coupling unit. This architecture is called “COSA+Coupling” for short. This type of finger has more flexible grasp between the distal phalanx and the middle phalanx. The coupling unit is simple in design. The finger that adopts “Coupling + COSA” is shown in Fig. 3b. It has opposite feature to “COSA+Coupling”. The “Under-actuation + COSA” and “COSA+Under-actuation” design are shown in Fig. 3c, 3d, respectively. Comparing to the former two designs, these two designs emphasize on adaptation more than human-like motion.

Finally, shown in Fig. 3e, “COSA+COSA” design is chosen for the three-phalanx finger in this paper, which means COSA is achieved on all the three phalanges. Therefore, comprehensive grasping ability is generated from the combination of coupling and self-adaptation. The finger has more grasp patterns that increase the possibility of successful grasp in unknown environments, though the design is the most complex one among the five possible designs addressed here. Geometric parameters are optimized to ensure the COSA motion on the three phalanges.

### 3.2 Mechanical Design

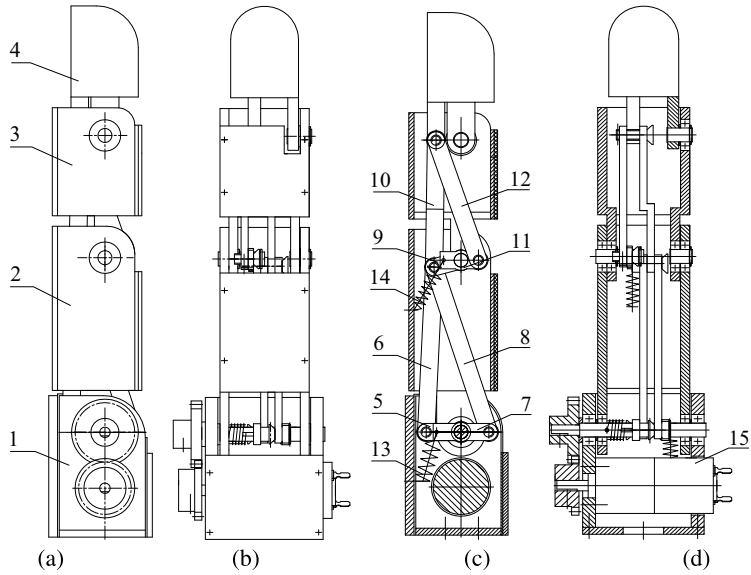
Mechanical design that can be manufactured is detailed in Fig. 4. There are two COSA units in the finger, the upper one and the lower one, actuated by one single motor. The driving bar, the first under-actuation bar, the connecting bar, the first coupling bar, the first changing bar, the proximal phalanx and the first decoupling spring compose the lower COSA unit.

The connecting bar, the second under-actuation bar, the distal phalanx, the second coupling bar, the second changing bar, the middle phalanx and the second decoupling spring compose the upper COSA unit. The first decoupling spring link the first changing bar to the base while the second decoupling spring link the second changing bar to the middle phalanx. The connecting bar transmits the motion from the lower COSA unit to the upper COSA unit. All bars are not exposed outside but covered by the surface of the finger. Thus the finger has humanoid appearance.

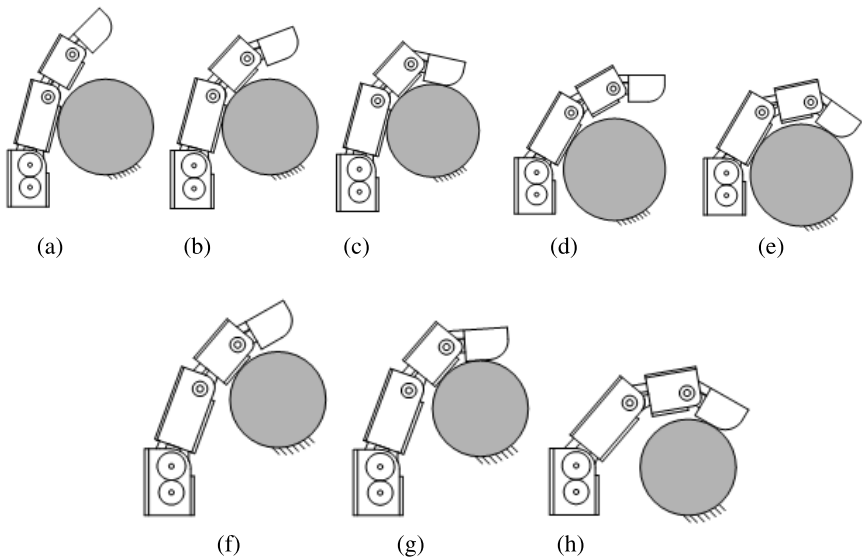
### 3.3 Movement and Grasping Patterns

The finger executes coupled motion before it contacts the object. The transmission ratio is dictated by the mechanical design. To simplify the analysis, the ratio of the middle phalanx rotating angle to the proximal phalanx rotating angle and the ratio of the distal phalanx rotating angle to the middle phalanx rotating angle are both set to be 1:1. The phalanges will rotate by the same angle about corresponding joints.

The grasping patterns can be classified according to which phalanges are restrained by the object. The first pattern is full-phalanx grasp shown in Fig. 5a-5c.



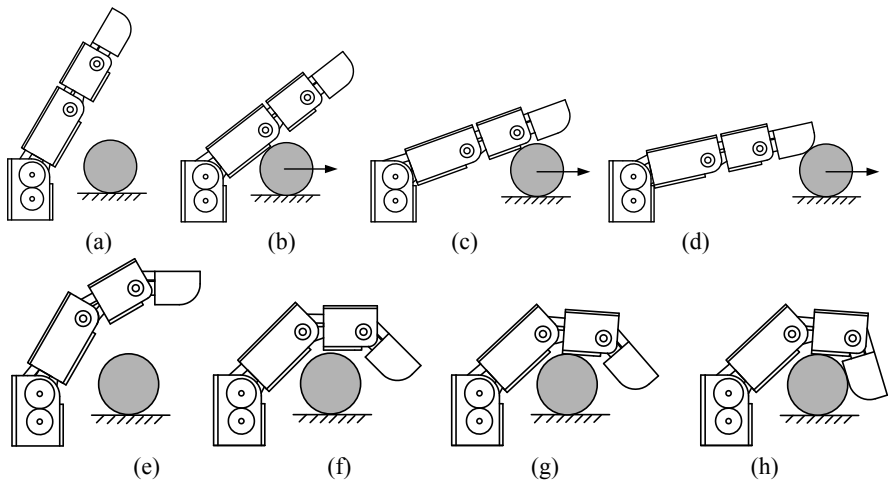
**Fig. 4.** Mechanical design of three-phalanx COSA finger. (a)Side view; (b)Front view; (c)Side cutaway view; (d)Front cutaway view. 1-Base; 2-Proximal phalanx; 3-Middle phalanx; 4-Distal phalanx; 5-Driving bar; 6-First under-actuation bar; 7-First changing bar; 8-First coupling bar; 9-Connecting bar; 10-Second under-actuation bar; 11-Second changing bar; 12-Second coupling bar; 13-First decoupling spring; 14-Second decoupling spring; 15-Motor.



**Fig. 5.** Different grasping patterns of three-phalanx COSA finger. (a)-(c)Full-phalanx grasp; (d), (e)Distal-proximal phalanx grasp; (f), (g)Distal-middle phalanx grasp; (h)Distal-phalanx grasp.

The proximal phalanx contacts the object first, and then the other two phalanges contact the object in series. The object is adaptively grasped by the finger. If the distal phalanx touches the object before the middle phalanx does but after the proximal phalanx due to the coupled motion, the object is grasped by only two phalanges shown in Fig. 5d, 5e. Another two-phalanx grasp is the distal-middle phalanx grasp in which the proximal phalanx does not contact the object shown in Fig. 5f, 5g. Finally, if the distal phalanx contact the object first and then blocked, the grasp is formed by only one phalanx shown in Fig. 5h. Note that the object in the figures is fixed. It can be seen that the four grasp patterns are dictated by the size, the shape and the position of the object. The multiple patterns enable the finger comprehensive grasping ability in contrast to traditional coupled fingers and self-adaptive under-actuated fingers. The advantage of COSA finger over a coupled finger is the self-adaptation apparently, which results in more stable grasps.

On the other hand, a comparison between COSA finger and traditional adaptive under-actuated finger is shown in Fig. 6 to illustrate the advantage of COSA over a traditional under-actuated finger.

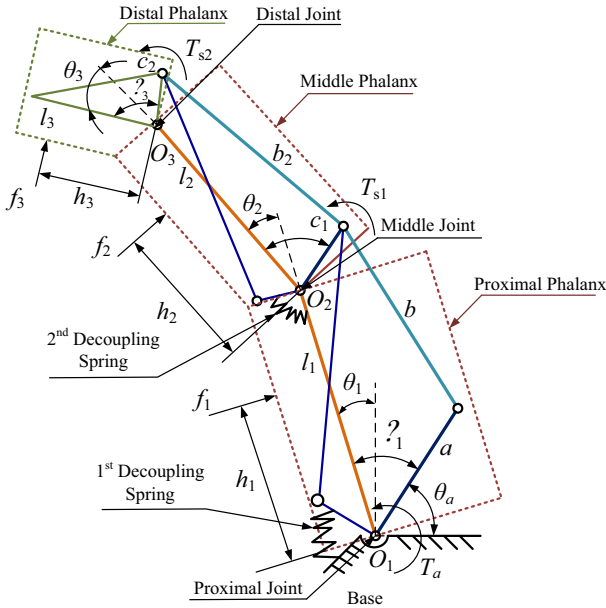


**Fig. 6.** Comparison between traditional under-actuated grasp and COSA grasp. (a)-(d)Traditional under-actuated grasp; (e)-(h)COSA grasp.

The cylinder to be grasped is put on a smooth surface so that the cylinder is able to slide easily. Figure 6a-6d show a traditional under-actuated finger bending to the object, and then pushing the object away instead of grasping it. This is highly likely on a smooth surface. The reason is that the traditional under-actuated finger executed adaptive grasp only if actuated by physical forces on the object. The upper phalanx will rotate towards the object when the lower phalanx is restrained by the object. If the object slides, there will be no force to actuate the finger thus the grasp will fail. The three-phalanx COSA finger is able to solve this problem by pre-shaping. As shown in Fig. 6e-6h, before any contact is made with the object, each phalanx has rotated by a certain angle to form an enclosure of the object. In another word, the finger pre-shapes. Finally, the object is grasped firmly.

### 4 Statics Analysis of Three-Phalanx COSA Finger

Static model is built on the three-phalanx finger architecture to obtain the grasping forces. The model is shown in Fig. 7.



**Fig. 7.** Static model of three-phalanx COSA finger, where  $O_1$ ,  $O_2$  and  $O_3$  are the centers of the joints, respectively,  $l_1$  is the distance between  $O_1$  and  $O_2$ ,  $l_2$  is the distance between  $O_2$  and  $O_3$ ,  $l_3$  is the length of the distal phalanx,  $a$ ,  $b_1$ ,  $c_1$ ,  $b_2$ , and  $c_2$  are the lengths of the corresponding bars,  $f_1$ ,  $f_2$  and  $f_3$  are the grasping force on the proximal phalanx, the middle phalanx and the distal phalanx, respectively,  $h_1$  is the force arm of  $f_1$  with respect to  $O_1$ ,  $h_2$  is the force arm of  $f_2$  with respect to  $O_2$ ,  $h_3$  is the force arm of  $f_3$  with respect to  $O_3$ ,  $\theta_a$  is the rotating angle of the driving bar with respect to  $O_1$ ,  $\theta_1$ ,  $\theta_2$  and  $\theta_3$  are the rotating angles of the proximal phalanx, the middle phalanx and the distal phalanx, respectively,  $\varphi_1$  is the angle between the driving bar and the proximal phalanx,  $\varphi_2$  is the angle between the connecting bar and the middle phalanx,  $\varphi_3$  is the inner angle of the distal phalanx, equaling ninety degrees here,  $T_{s1}$  is the torque of the first decoupling spring,  $T_{s2}$  is the torque of the second decoupling spring,  $T_a$  is the torque of the motor.

According to the principle of virtual work, one obtains

$$\mathbf{t}^T \boldsymbol{\omega}_a = \mathbf{f}^T \mathbf{V}, \tag{1}$$

where  $\mathbf{t}$  is the input torque vector exerted by the motor and the under-actuation spring,  $\mathbf{f}$  is the grasping force vector on the two phalanges,  $\boldsymbol{\omega}_a$  is the corresponding rotating velocity vector,  $\mathbf{V}$  is the velocity vector of contact points.

The velocity vector is related to the rotating angles and corresponding arms, therefore it is expressed as:

$$\mathbf{V} = \mathbf{J}_v \dot{\boldsymbol{\theta}} = \begin{bmatrix} h_1 & 0 & 0 \\ l_1 \cos \theta_2 + h_2 & h_2 & 0 \\ l_1 \cos(\theta_2 + \theta_3) + l_2 \cos \theta_3 + h_3 & l_2 \cos \theta_3 + h_3 & h_3 \end{bmatrix} \begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \\ \dot{\theta}_3 \end{bmatrix}, \tag{2}$$

in which  $\mathbf{J}_v$  is expressed as:

$$\mathbf{J}_v = \begin{bmatrix} h_1 & 0 & 0 \\ l_1 \cos \theta_2 + h_2 & h_2 & 0 \\ l_1 \cos(\theta_2 + \theta_3) + l_2 \cos \theta_3 + h_3 & l_2 \cos \theta_3 + h_3 & h_3 \end{bmatrix}. \tag{3}$$

According the relationship among bars in four-bar linkage, it is obtained that

$$\begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \\ \dot{\theta}_3 \end{bmatrix} = \mathbf{J}_\omega \boldsymbol{\omega}_a = \begin{bmatrix} 1 & A & -AB \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{\theta}_a \\ \dot{\theta}_2 \\ \dot{\theta}_3 \end{bmatrix}, \tag{4}$$

where

$$\mathbf{J}_\omega = \begin{bmatrix} 1 & A & -AB \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{5}$$

$$A = \frac{c_1[l_1 \sin(\theta_2 - \phi_2) - a \sin(\phi_1 - \phi_2 + \theta_2)]}{a[l_1 \sin(\phi_1) + c_1 \sin(\phi_1 - \phi_2 + \theta_2)]}, \tag{6}$$

$$B = \frac{c_2[l_2 \sin(\theta_3 - \phi_3) - c_1 \sin(\phi_2 - \phi_3 + \theta_3)]}{c_1[l_2 \sin(\phi_2) + c_2 \sin(\phi_2 - \phi_3 + \theta_3)]}. \tag{7}$$

Finally by solving eq. (1), one gets

$$\mathbf{f}^T = \mathbf{t}^T \mathbf{J}_\omega^{-1} \mathbf{J}_v^{-1}, \tag{8}$$

$$\mathbf{f} = \begin{bmatrix} \frac{T_a}{h_1} + \frac{AXT_a}{h_1 h_2} + \frac{ABXYT_a}{h_1 h_2 h_3} - \frac{ABZT_a}{h_1 h_3} - \frac{XT_{s1}}{h_1 h_2} - \frac{ZT_{s2}}{h_1 h_3} \\ -\frac{AT_a}{h_2} - \frac{ABYT_a}{h_2 h_3} + \frac{T_{s1}}{h_2} - \frac{YT_{s2}}{h_2 h_3} \\ \frac{ABT_a}{h_3} + \frac{T_{s2}}{h_3} \end{bmatrix}. \tag{9}$$

The torques exerted by the springs is ignored, since their magnitudes are relative small comparing to the motor torque. It has been verified by practice this assumption will not influence the effect of static analysis. Equation (9) is reformed without spring torques as:

$$\mathbf{f} = \begin{bmatrix} \frac{h_2 h_3 + AXh_3 + ABXY - ABZ}{h_1 h_2 h_3} T_a \\ -\frac{Ah_3 + ABY}{h_2 h_3} T_a \\ \frac{AB}{h_3} T_a \end{bmatrix}. \tag{10}$$

Equation (10) gives a simple result in which grasping forces on the three phalanges are linear functions of the motor torque. In above equation, it is noted that

$$X = l_1 \cos \theta_2 + h_2, \tag{11}$$

$$Y = l_2 \cos \theta_3 + h_3, \tag{12}$$

$$Z = l_1 \cos(\theta_2 + \theta_3) + l_2 \cos \theta_3 + h_3. \tag{13}$$

Based on the coupled motion design, the rotating angles justify the following relationship

$$\theta_3 \geq \theta_2 \geq \theta_1. \tag{14}$$

The grasping forces obtained from are primarily dictated by two factors. One is the configuration of the finger, described by  $\theta_2$  and  $\theta_3$ . The other factor is the contact locations on the phalanges, described by force arms  $h_1$ ,  $h_2$  and  $h_3$ . Geometric parameters for calculating grasping forces are shown in Table 1. To reduce the number of variables,  $h_1$ ,  $h_2$  and  $h_3$  are given by 15, 20 and 25 mm, respectively. Hence, each force is expressed as a function of  $\theta_2$  and  $\theta_3$ . The forces are drawn in Fig. 8.

**Table 1.** Geometric parameters of finger

$a$	$b_1$	$c_1$	$b_2$	$c_2$	$l_1$	$l_2$	$\phi_3$
10	52	8	42	7	50	35	90°

It should be remembered that only half of the plane formed by  $\theta_2$  and  $\theta_3$  are shown in the figure. The grasping forces are generally larger than tendon-based robot fingers. The maximum of  $f_1$ ,  $f_2$  and  $f_3$  reaches about 66.7 N, 883.6 N and 152.1 N, respectively. The large output is the typical feature of linkage mechanisms.

From the figure it can be seen that  $\theta_3$  impacts the force magnitudes more distinctly than  $\theta_2$  does. Increasing of  $\theta_3$  leads to decreasing of  $f_1$  and  $f_3$  but increment of  $f_2$ . The effect of  $\theta_2$  is complex. When  $\theta_2$  declines, both  $f_2$  and  $f_3$  rise slightly.  $f_1$  will decrease if  $\theta_3$  is more than about sixty-five while it will increase if  $\theta_3$  is less than about sixty-five. By finding the planes where the forces equal to zero, the stability of grasp is shown in Fig. 9. A stable grasp refers to a positive force on the phalanx.

If the force is negative, the corresponding phalanx might slide on the object or lose contact with the object. There are two unstable areas and one stable area in Fig. 9. The first unstable area emerges when  $\theta_2$  is less than about sixty degrees and  $\theta_3$  is more than about seventy degrees in the meantime. In this area,  $f_1$  is negative and the other two forces are positive. The second unstable area exists where  $\theta_3$  is less than about sixty-two degrees.  $f_2$  is the only negative force in this area.

The other part is the stable area in which all three forces are positive. Although this situation is ideal in the grasp, it does not mean the finger is only able to grasp in this stable area. In most cases of a multi-fingered grasp, only force on the terminal phalanx i.e.  $f_3$  needs to be positive. As it shows in the figure,  $f_3$  is always positive. Therefore, the three-phalanx COSA finger has good grasp stability.

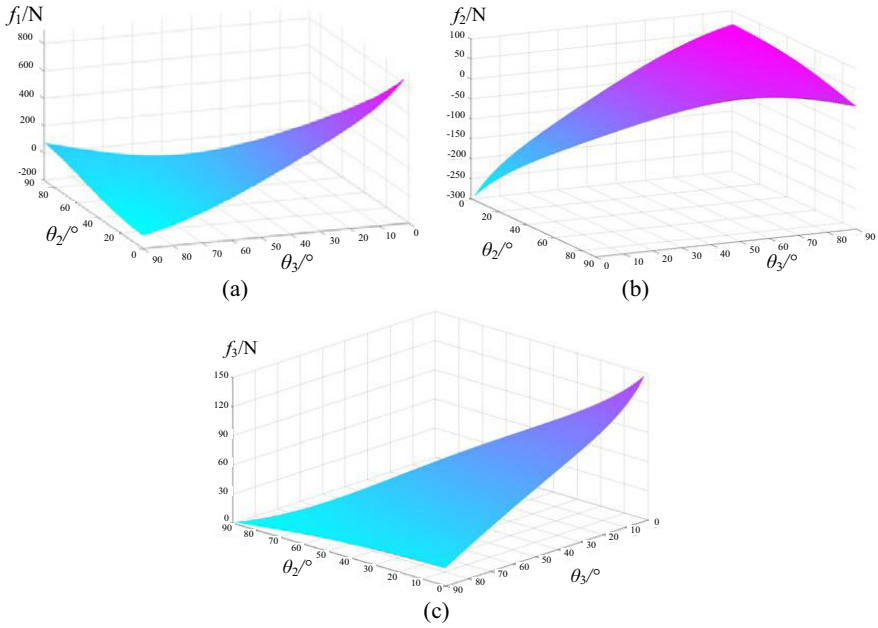


Fig. 8. Grasping forces of three-phalanx COSA finger (a)  $f_1$ ; (b)  $f_2$ ; (c)  $f_3$ .

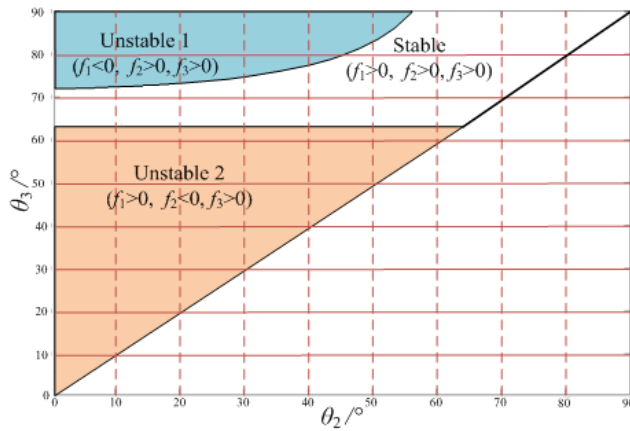


Fig. 9. Grasping stability of the three-phalanx COSA finger

## 5 Conclusion

In this paper, the design and study of three-phalanx COSA finger was addressed. The principle of COSA was proposed as a third approach of under-actuation apart from

rigid coupling and self-adaptation. Linkage COSA unit used in two-phalanx COSA finger is the foundation of three-phalanx finger design.

Mechanical design was developed after discussion on geometric designs. The advantage in COSA grasping was illustrated especially by a comparison with traditional under-actuated fingers.

The finger has good grasp stability verified by force simulation. In summary, a simple robotic finger with humanoid appearance, human-like motion and comprehensive grasping capability is developed in this paper. The advantages of the COSA finger make it fit the requirement of either prosthetic hands or humanoid robot hands.

**Acknowledgements.** This research was supported by National Natural Science Foundation of China (No. 50905093) and Hi-Tech R&D Program of China (No. 2007AA04Z258).

## References

1. Laliberte, T., Gosselin, C.: Simulation and Design of Under-actuated Mechanical Hands. *Mech. Mach. Theory* 33(1/2), 39–57 (1998)
2. Pons, J.L., Ceres, R., Pfeiffer, F.: Multifingered Dexterous Robotics Hand Design and Control: A Review. *Robotica*, 661–674 (1999)
3. Pons, J.L., Rocon, E., Ceres, R., et al.: The MANUS-HAND\* Dexterous Robotics Upper Limb Prosthesis: Mechanical and Manipulation Aspects. *Autonomous Robots* 16, 143–163 (2004)
4. Dollar, A.M., Howe, R.D.: Towards Grasping in Unstructured Environments: Grasper Compliance and Configuration Optimization. *Advanced Robotics* 19(5), 523–543 (2005)
5. Dollar, A.M., Howe, R.D.: The SDM Hand as a Prosthetic Terminal Device: A Feasibility Study. In: *IEEE 10th Int. Conf. on Rehabilitation Rob.*, Noordwijk, Netherlands, pp. 978–983 (2007)
6. Dechev, N., Cleghorn, W.L., Naumann, S.: Multiple Finger, Passive Adaptive Grasp Prost. *Mech. Mach. Theory* 36, 1157–1173 (2001)
7. Laliberte, T., Birglen, L., Gosselin, C.: Under-actuation Robotic Grasping Hands. *Machine Intelligence & Robotic Control* 4(3), 1–11 (2002)
8. Birglen, L., Gosselin, C.M.: Kinetostatic Analysis of Underactuated Fingers. *IEEE Trans. on Rob. and Aut.* 20(2), 211–221 (2004)
9. Birglen, L., Birglen, L., Gosselin, C.M.: Geometric Design of Three-Phalanx Underactuated Fingers. *ASME Journal of Mechanical Design* 128(2), 356–364 (2006)
10. Zhang, W., Che, D., Liu, H., et al.: Super Under-actuated Multi-fingered Mechanical Hand with Modular Self-adaptive Gear-rack Mechanism. *Industrial Robot: An International Journal* 36(3), 255–262 (2009)
11. Zhang, W., Che, D., Chen, Q., et al.: Study on Gesture-Changeable Under-actuated Humanoid Robotic Finger. *Chinese Journal of Mechanical Engineering* 23(2), 142–148 (2010)
12. Wu, L., Ceccarelli, M.: A Numerical Simulation for Design and Operation of an Under-actuated Finger Mechanism for LARM Hand. *Mechanics Based Design of Structure and Machines* 37(1), 86–112 (2009)
13. Wu, L., Carbone, G., Ceccarelli, M.: Designing an Underactuated Mechanism for a 1 Active DOF Finger Operation. *J. Mech. Mach. Theory* 44(2), 336–348 (2009)
14. Yang, D., Zhao, J., Gu, Y., et al.: An Anthropomorphic Hand Developed Based on Under-actuated Mechanism and Controlled by EMG Signals. *Journal of Bionic Engineering*, 255–263 (2009)



# Design of a Bionic Saltatorial Leg for Jumping Mini Robot

Xin Fu<sup>1</sup>, Fei Li<sup>1</sup>, Weiting Liu<sup>1,\*</sup>, Cesare Stefanini<sup>2</sup>, and Paolo Dario<sup>2</sup>

<sup>1</sup> The State Key Lab of Fluid Power Transmission and Control, Zhejiang University,  
310027 Hangzhou, P.R. China  
{xfu, li\_fei, liuwt}@zju.edu.cn

<sup>2</sup> CRIM Lab, Polo Sant'Anna Valdera, Viale Rinaldo Piaggio 34,  
56025 Pontedera (Pisa), Italy  
{cesare, dario}@sssup.it

**Abstract.** This paper presents a bio-inspired design of an artificial saltatorial leg for a jumping mini robot. By using a rebuilt traditional four-bar linkage mechanism, the artificial saltatorial leg is able to successfully imitate the characteristic of a real jumping insect, kinematically and dynamically. According to the simulation analysis, jumping performance is improved by choosing a better tibia/femur ratio and a proper mass center placement. By means of high speed camera experiment, the jumping characteristic among insect, dynamic model and robot is compared; the dynamic similarity between insect and the robot is shown at last.

**Keywords:** Artificial Saltatorial Leg, Jumping Robot, Bionic Robot, Four-bar Linkage Mechanism, Bionics.

## 1 Introduction

For animals in the natural world, jumping is more than a method to overcome obstacles higher than themselves; it also improves their locomotion ability and offers a way to escape from emergency. As one of the most efficient natural locomotion modes, jumping is preferred by a lot of animals from mammals, amphibians to insects (Fig. 1), especially when animals travel on tough terrains or for those who benefit from their scale effects [1].

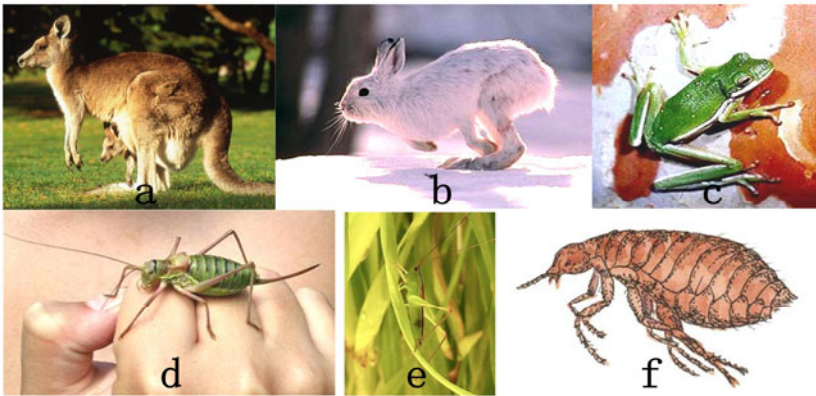
During the past decades, based on the biological knowledge learned from the natural world, various robot prototypes have been built in order to find a better solution for robot locomotion besides wheels and crawler, as well as to improve the efficiency and stability of the whole locomotion system. For example, a simple but efficient leg-wheeled locomotion developed based on the traditional wheel structure offers good passing ability for robots [2-4]; walking locomotion inspired from mammals provides robots low energy consumption and better moving stability on tough terrains despite a

---

\* Corresponding author.

complex controlling strategy and kinematic complication [5-9]; novel jumping robots inspired from kangaroo [10] and fleas [11-12] are also developed and proves the feasibility of using jumping as the primary moving mode for robots.

Based on above lessons, our robot prototype GRILLO, a millimeter-sized mobile robot prototype with onboard power supply integrated and designed as a mobile roboticized platform for environmental monitoring, rescue or exploration in unstructured environments, is presented in this paper. By imitating and optimizing the catapult jumping adopted by insects, this jumping robot offers a lot of advantages such as its good moving ability and better practicability.



**Fig. 1.** Jumping animals in the natural world: (a) kangaroo; (b) rabbit; (c) frog; (d) cricket; (e) leafhopper; (f) flea

## 2 Saltatorial Leg Design

### 2.1 Four-Bar Linkage Mechanism Model

According to some previous research, two jumping styles including countermovement jumping and catapult jumping are usually adopted by vertebrates (e.g. human) and insects (e.g. flea, froghopper and locust, etc.) respectively [13-14]. For those jumping insects, almost all energy used to generate a jump is coming from the simultaneous extension of their two long hind legs (saltatorial legs). As an example, by extending the tarsus along a straight line during leg elongation, the extensor of a 1.5-2.0 g weight locust (*Schistocerca gregaria*) is able to release 9-11 mJ energy in 30 ms thus lead to a take off velocity of about 3.2 m/s [15].

Therefore, considering its simple design and high efficiency, the catapult design is more applicable for our jumping robot. In order to imitating the jumping kinetics of a natural saltatorial leg, a planar four-bar linkage mechanism as shown in fig. 2 is adopted by GRILLO firstly.

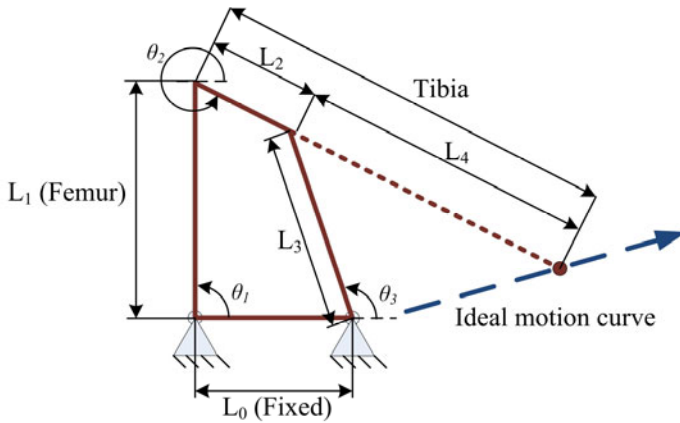


Fig. 2. Planar four-bar-linkage mechanism model

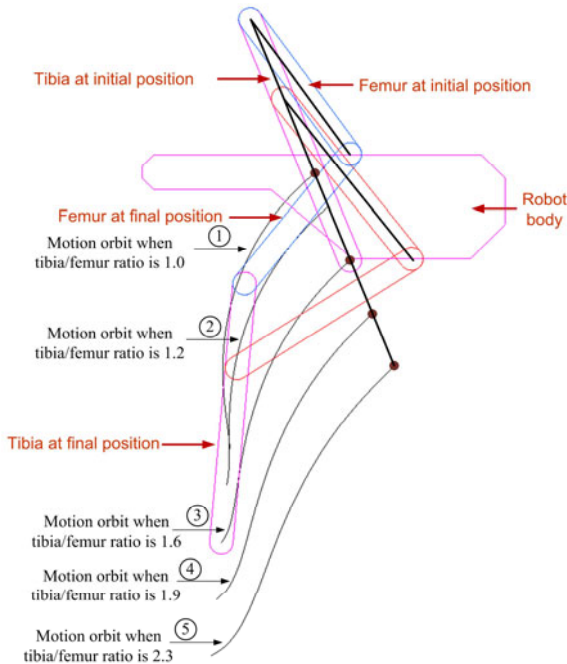


Fig. 3. The applicable design and approximately linear elongation orbits with different tibia/femur ratio. It indicates that along with the prolongation of tibia bar, orbits become more similar and the final motion orbit only depends on the tibia/femur ratio.

The simplified planar four-bar linkage model used to achieve a linear motion at the tip of tibia is composed of five bars, including a fixed bar  $L_0$ , the femur  $L_1$ , the tibia  $L_2$  and  $L_4$ , as well as an auxiliary bar  $L_3$ . Such a structure can be analyzed based on the following mathematical model,

$$\begin{bmatrix} L_2 \sin \theta_2 - L_3 \sin \theta_3 \\ L_2 \cos \theta_2 - L_3 \cos \theta_3 \end{bmatrix} \begin{bmatrix} \omega_2 \\ \omega_3 \end{bmatrix} = \begin{bmatrix} -L_1 \sin \theta_1 \\ -L_1 \cos \theta_1 \end{bmatrix} \cdot \omega_1 \tag{1}$$

$$\begin{aligned} \theta_{10} &= \frac{\pi}{2} \\ \theta_{20} &= \arccos\left(\frac{L_0^2 + L_1^2 + L_2^2 - L_3^2}{2L_2\sqrt{L_0^2 + L_1^2}}\right) - \arctan\left(\frac{L_1}{L_0}\right) \\ \theta_{30} &= \pi - \arccos\left(\frac{L_0^2 + L_1^2 + L_3^2 - L_2^2}{2L_3\sqrt{L_0^2 + L_1^2}}\right) - \arctan\left(\frac{L_1}{L_0}\right) \end{aligned} \tag{2}$$

where  $\omega_i$  and  $L_i$  ( $i=0, 1, 2, 3, 4$ ) is the angular velocity and the length of each bar respectively.  $\theta_1, \theta_2, \theta_3$  are the initial angular value of  $L_1, L_2$  and  $L_3$ . Considering the structure compact and assembling convenience, different relative value among  $L_1, L_2$ , and  $L_3$  are tested, then an applicable design and approximately linear elongation orbits of the tarsus shown in fig. 3 are achieved. For the tested model, the fixed bar ( $L_0$ ) is 21mm long, the femur ( $L_1$ ) is 29mm long, and upper part of tibia ( $L_2$ ) is 15mm long while the lower part of tibia ( $L_4$ ) is ranging from 14mm to 50mm to gain different femur/tibia ratio from 1.0 to 2.3, which are selected according to our biological observation (e.g. for human, the ratio is about 1.0, for leafhopper, the ratio is about 1.6).

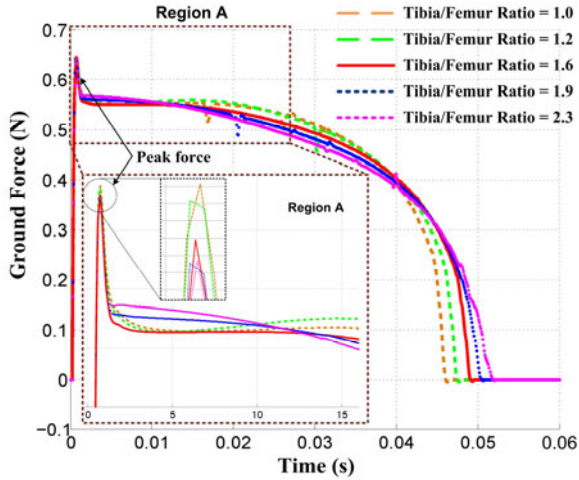
### 2.2 Dynamic Optimization

Dynamic analysis is then carried out based on the given four-bar linkage structure mentioned above. The ratio between leg mass and body mass is set to be 1/10, 1 gram and 10 gram respectively. Motion characteristics only depend on the tibia/femur ratio defined as following:

$$tibia/femur\_ratio = \frac{L_2 + L_4}{L_1} \tag{3}$$

The above kinematic analysis is aimed to find an applicable design and approximately linear elongation for the saltatorial leg. Along with the increasing of tibia bar length, orbits become more similar (e.g. orbit 4 & 5). Then in order to find an optimized tibia/femur ratio to minimize the stress of the leg structure, dynamic simulation of such a jumping mechanism is run.

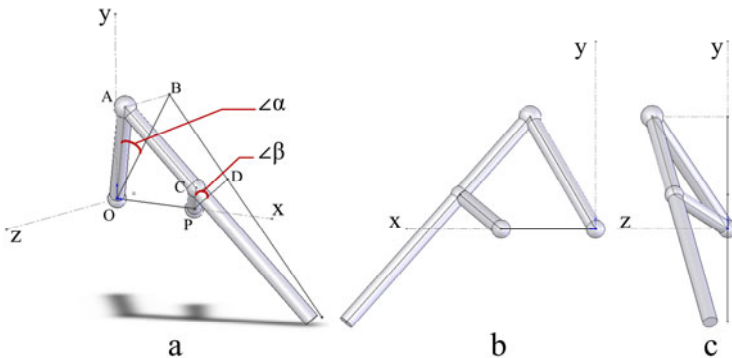
As shown in fig. 4, five different tibia/femur ratios from 1.0 to 2.3 based on the leg structure in fig. 3 are tested respectively with ADAMS simulation, and the force exerted on the ground during jumping acceleration is measured. The peak force of each curve exists 1-2 ms after jumping. Such a sudden rise in force may result in a rupture of the skeleton structure or slippage between tarsus and ground. According to our simulation, this peak force decreases with increasing of the tibia/femur ratio. On the other hand, the maximum ground force is not monotonic decreasing after the spike and the minimum one exits when tibia/femur ratio is around 1.6 (the red solid curve). Considering the engineering requirements (e.g. less influence of the short time peak force, long time stability after the spike, structure design, etc), among the five tested ones, a tibia/femur ratio around 1.6 (red solid curve) is regarded as a better one because of the moderate peak force and the minimized ground force after the spike.



**Fig. 4.** The dynamic characteristics during jumping acceleration of different leg designs (tibia/femur ratios from 1.0 to 2.3). The peak force of each curve exists 1-2 ms after jumping and the maximum ground force after this spike exits when tibia/femur ratio is around 1.6 (the red solid curve).

### 2.3 Biological Approximation

To develop the biological approximation of this four-bar linkage mechanism, as shown in fig. 5, planar structure is then rebuilt to a spatial structure by rotating  $L_1$  and  $L_3$  along x-axis. The rotating angle  $\alpha$  is larger than angle  $\beta$  so the tarsus are able to be positioned directly beneath the body and pressed against each other during jumping



**Fig. 5.** Structure rebuilding of the saltatorial leg design for biological approximation. The rotating angle  $\alpha$  is larger angle  $\beta$  so the tarsus are able to be positioned directly beneath the body and pressed against each other during jumping acceleration. Although such a design increases the complexity and introduces more joints, it reduces the chance of asymmetric thrust forces between two legs, as well as the destabilizing rotation during the flight phase.

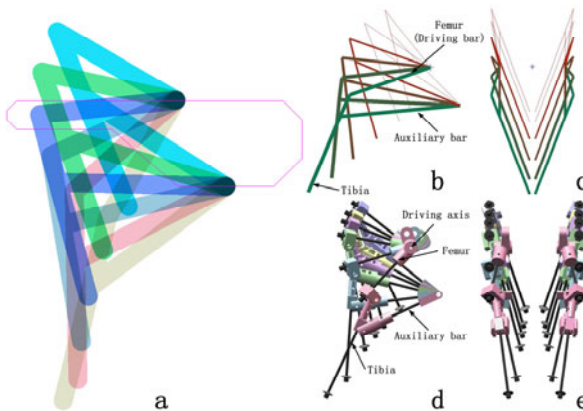
acceleration. One should notice that such a rebuilding will not influence on the kinematics characteristics of the previous planar design when consider the plane OAB as a femur and the plane PCD as the auxiliary bar. Although such a design increases the complexity and introduces more joints, it reduces the chance of asymmetric thrust forces between two legs, as well as the destabilizing rotation during the flight phase.

Considering the structural size of the final fabricated robot body, a tibia/femur ratio of 1:1.72 is finally adopted. Such a value is very close to the biological one according to the statistical data from our investigation on leafhoppers [16]. From a biological viewpoint, this may also represent an evolutionary optimization of this specific species that develop their catapult jumping ability.

### 3 Jumping Robot Fabrication

#### 3.1 Fabrication of Artificial Saltatorial Legs

Fig. 6 shows the motion curve comparison between models and the fabricated artificial saltatorial legs. The legs are able to elongate about 27.4mm, almost equal to its 27mm long tibia (an  $80^\circ$  rotation angle for the femur). The actual extension motions between the ideal model and the fabricated legs are different mainly because of the machining difficulties of millimeter sized sphere joint, of which we use three respective rotating joints instead. Furthermore, the jumping energy is coming from a continuous rotating DC motor, a segment-gear mechanism including a spring connecting between femur and robot body as an energy storage element is used to imitate the natural muscle-tendon system as the actuating device of the prototype [16]. It drives the legs and transmits the motor's continuous rotation to a reciprocal load-release motion.



**Fig. 6.** The comparison of motion characteristics between models and the fabricated artificial saltatorial legs. The actual extension motions between the ideal model and the fabricated legs are different mainly because of the machining difficulties of millimeter sized sphere joint, of which we use three respective rotating joints instead. (a) Motion characteristic of planar saltatorial legs; (b & c) motion characteristic of spatial saltatorial legs; (d & e) motion characteristic of the finally fabricated saltatorial legs.

### 3.2 Mass Centre Readjustment

For a robot regarding jumping as its primary locomotion mode, the moving stability (e.g. jumping stability, flight stability and landing stability) is one of the key issues in the prototype design. The influence of the mass center position on jumping stability is analyzed firstly. Fig. 7 indicates that when the mass center of the whole body is placed in the ideal zone, the insect jumps with a little anticlockwise will not influence on the jumping stability. And it improves the flight and landing stability which is also significant for our robot design.

Unfortunately, it is unpractical to confirm the precise position of the mass center before assembling due to its structural complication. So the robot is assembled firstly, and then readjusted by adding extra balance weight on the body for jumping stability improvement. The flight and landing stability are further improved respectively by adding wings and a tail on the body, as well as an optimization of the passive forelegs through computational simulation. So compared with the one before readjustment, the

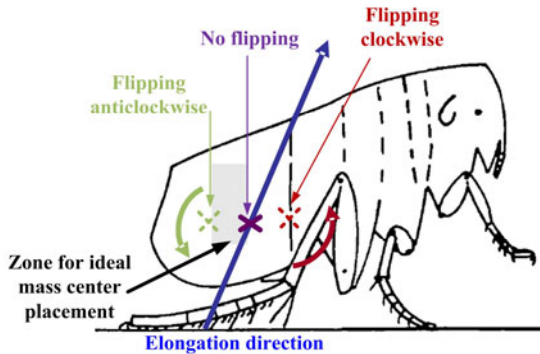


Fig. 7. Analysis of the jumping stability influenced by the mass center position [17]

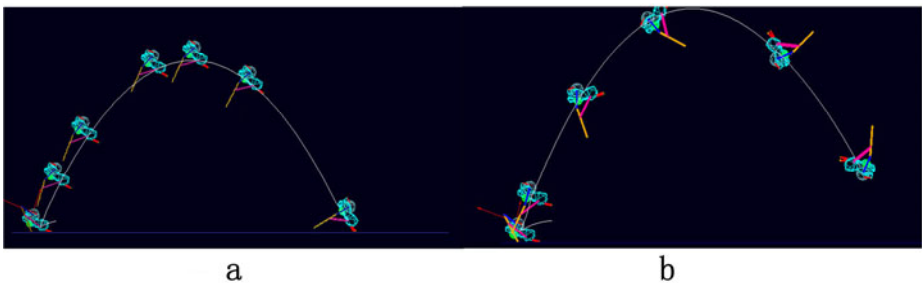
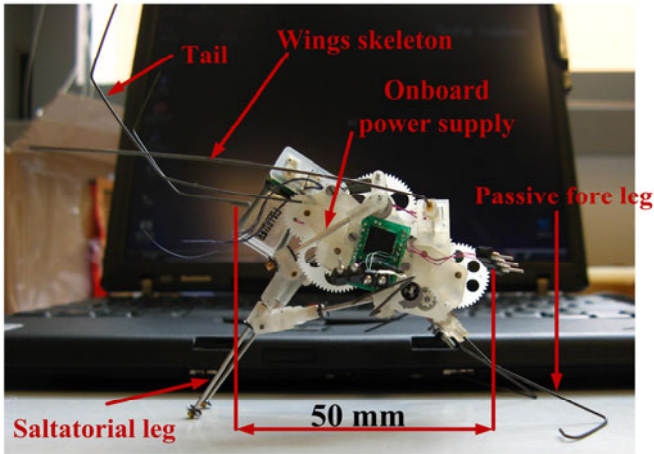


Fig. 8. Improvement result of the mass center readjustment. Because of the structural complication, the jumping stability of the prototype is readjusted by adding extra balance weight on the body after assembling. In a meanwhile, the flight and landing stability are also improved respectively by adding wings and a tail on the body, as well as an optimization of the passive forelegs. Optimizations are analyzed with ADAMS simulation.

finally assembled jumping robot prototype GRILLO III shows a better continuous jumping performance. Fig. 8 shows the improvement result of the mass center readjustment by ADAMS simulation analysis. Fig. 9 shows the finally fabricated jumping robot GRILLO III. It is a jumping robot with 50mm x 20mm x 25mm (L x W x H) in dimension and 22 g in weight including the onboard power supply.



**Fig. 9.** Jumping robot prototype GRILLO III is the improved edition of its predecessor [18, 16]. Right now it is a jumping robot with 50mm x 20mm x 25mm (L x W x H) in dimension and 22g in weight including the onboard power supply.

## 4 Experiments and Comparison

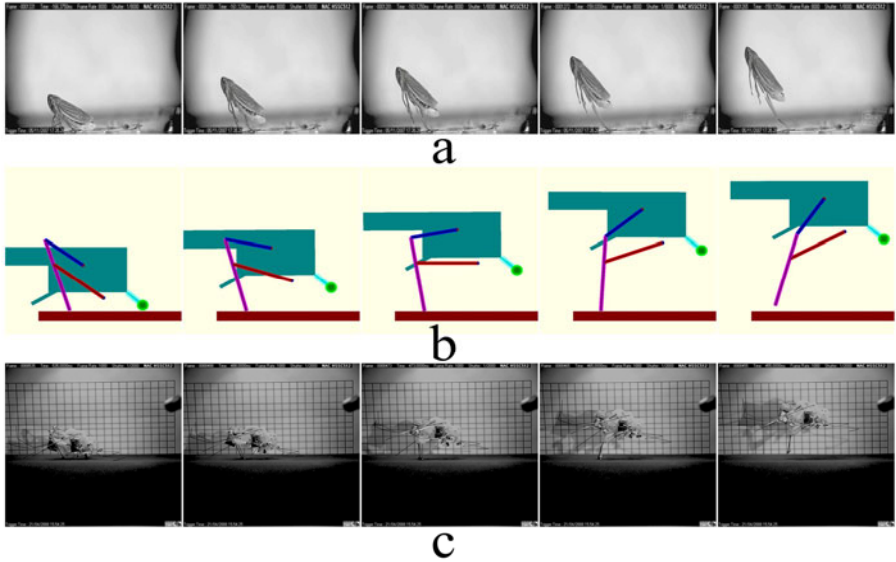
According to some biological observation by other research, three distinct phases are observed with the jumping insect leafhopper from preparing to the final launching [19]. The first one is a leivation phase of about 15-30 ms, in which the insect positions its hind legs directly beneath the body with tarsus pressed against each other. Then it is the holding phase of about 10-200 ms, in which the hind legs remain and the take-off angle is adjusted by front and middle legs. At last the jump is generated by simultaneous extension of the two long hind legs. This rapid jumping phase lasts 5-6 ms.

Fig. 10 shows the comparison among leafhopper jumping [16], jumping simulation and high speed snapshot of the jumping robot prototype. At the beginning, the robot stands on the ground, and its actuating system starts to load the spring to restore enough energy for its jumping. After loading, the fast releasing mechanism is then activated and the femur rotates anticlockwise, the saltatorial legs elongate rapidly to force the robot jump as a result of spring shortening.

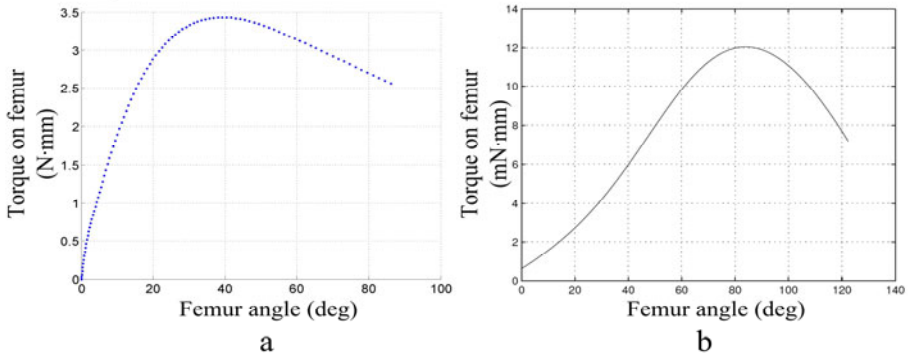
According to our simulation and experiment results, the elongation usually costs about 30ms. The jumping acceleration is about  $200 \text{ m/s}^2$  for an ideal model (ignoring friction), and about  $59 \text{ m/s}^2$  for our fabricated prototype.

Fig. 11 shows the dynamic similarity between the jumping robot and the leafhopper we measured in our biology observations [1].





**Fig. 10.** The comparison among leafhopper jumping (a), jumping simulation (b) and high speed snapshot of the jumping robot (c)



**Fig. 11.** Dynamic similarity between the jumping robot (a) and the leafhopper (b) we measured in our biology observations [1]

## 5 Conclusions and Future Work

By imitating the natural saltatorial legs of jumping insects, a bio-inspired artificial saltatorial leg for jumping mini robot is designed and tested. Such a saltatorial leg is designed by adopting a traditional four-bar linkage mechanism and then rebuilding it to a spatial one in order to reproduce the jumping characteristic of a real jumping insect both kinematically and dynamically.

According to the simulation analysis, with a better femur/tibia ratio close to the natural one, the jumping stability can be improved because of the moderate peak force after acceleration beginning and the minimized maximum ground force after this spike. Thus the risk of skeleton rupture and tarsus slippage can be reduced.

Jumping robot GRILLO is then fabricated based on this leg design. By means of mass center readjustment, the moving stability is improved. With high speed camera experiment, the jumping characteristic among insect, dynamic model and robot are compared. The experiment result shows the dynamic similarity between insect and the robot.

Future work will be focused on the feasibility of the sensing and controlling system (e.g. hybrid control method [20]). The sensing system is based on artificial hair receptor which is also inspired from natural world. Integrating aliened piezoelectric fibers with different aspect ratio on the body provides the robot with sensing ability to flow variation, pressure changing or vibration in different frequency. The controlling system is used to improve the jumping efficiency by autonomous adjustment of jumping angle based on the information from above sensing system.

**Acknowledgments.** The work presented in this paper is support by International S&T Cooperation Program of China (No. 2008DFR70410) and Zhejiang Provincial Natural Science Foundation of China (R105008) respectively. Authors also would like to thank Dr. Umberto Scarfogliero and Dr. Gabriella Bonsignori of Scuola Supreiore Sant'Anna, Italy, for their valuable advices on the mechanical design as well as the outstanding work on the jumping locomotion investigation on leafhoppers.

## References

1. Scarfogliero, U., Bonsignori, G., Stefanini, C., Sinibaldi, E., et al.: Bioinspired Jumping Locomotion in Small Robots: Natural Observation, Design, Experiments. Springer. Tracts. In. *Adv. Robotics* 54, 329–338 (2009)
2. Birch, M., Quinn, R., Hahm, G., Phillips, S.M., et al.: A Miniature Hybrid Robot Propelled by Legs. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Maui, USA (2001)
3. Lambrecht, B.G.A., Horchler, A.D., Quinn, R.D.: A Small, Insect-Inspired Robot that Runs and Jumps. In: *IEEE International Conference on Robotics and Automation*, Barcelona, Spain (2005)
4. Saranli, U.: RHex: A Simple and Highly Mobile Hexapod Robot. *Int. J. Robot. Res.* 20, 616–631 (2001)
5. Clark, J.E., Cham, J.G., Bailey, S.A., Froehlich, E.M., et al.: Biomimetic Design and Fabrication of a Hexapedal Running Robot. In: *IEEE International Conference on Robotics and Automation*, Seoul, Korea (2001)
6. Nelson, G.M., Quinn, R.D., Bachmann, R.J., Flannigan, W.C., et al.: Design and Simulation of a Cockroach-like Hexapod Robot. In: *IEEE International Conference on Robotics and Automation*, Albuquerque, USA (1997)
7. Berns, K., Ilg, W., Deck, M., Albiez, J., et al.: Mechanical Construction and Computer Architecture of the Four-legged Walking Machine BISAM. *IEEE/ASME Trans. Mechatron.* 4, 32–38 (1999)

8. Kimura, H.: Adaptive Dynamic Walking of a Quadruped Robot on Natural Ground Based on Biological Concepts. *Int. J. Robot. Res.* 26, 475–490 (2007)
9. Geyerl, H., Seyfarthl, A., Blickhan, R.: Compliant Leg Behaviour Explains Basic Dynamics of Walking and Running. *Proc. R. Soc. B* 273, 2861–2867 (2006)
10. Hyon, S.H., Mita, T.: Development of a biologically inspired hopping robot- “Kenken”. In: *IEEE International Conference on Robotics and Automation*, pp. 3984–3991. IEEE Press, Washington (2002)
11. Kovac, M., Fuchs, M., Guignard, A., Zufferey, J.C., et al.: A Miniature 7g Jumping Robot. In: *IEEE International Conference on Robotics and Automation*, pp. 373–378. IEEE Press, Pasadena (2008)
12. Kovac, M., Schlegel, M., Zufferey, J.C., Floreano, D.: Steerable miniature jumping robot. *Auton. Robot.* 28, 295–306 (2010)
13. Komi, P.V., Bosco, C.: Utilization of Stored Elastic Energy in Leg Extensor Muscles by Men and Momen. *Med. Sci. Sport. Exer.* 10, 261–265 (1978)
14. Bennet-Clark, H.C.: *The insect integument*. Elsevier Science & Technology, Amsterdam (1976)
15. Bennet-Clark, H.C.: The energetics of the jump of the locust *Schistocerca gregaria*. *J. Exp. Biol.* 63, 53–83 (1975)
16. Li, F., Bonsignori, G., Scarfogliero, U., Chen, D., et al.: Jumping Mini-robot with Bio-inspired Legs. In: *Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics*, pp. 933–938. IEEE Press, Bangkok (2009)
17. Bennet-Clark, H.C., Lucey, E.C.A.: The Jump of the Flea - A Study of the Energetics and a Model of the Mechanism. *J. Exp. Biol.* 47, 59–76 (1967)
18. Scarfogliero, U., Stefanini, C., Dario, P.: A Bioinspired Concept for High Efficiency Locomotion in Micro Robots: The Jumping Robot Grillo. In: *IEEE International Conference on Robotics and Biomimetics*, Kunming, China (2006)
19. Burrows, M.: Kinematics of jumping in leafhopper insects (Hemiptera, Auchenorrhyncha, Cicadellidae). *J. Exp. Biol.* 210, 3579–3589 (2007)
20. Li, Y.M.: A Hopping Robot Control Using Hybrid Control Methodology. In: *Proceedings of World Multiconference on Systemics, Cybernetics and Informatics (SCI 1998)*, Orlando, pp. 77–82 (1998)

# Asymmetrical Prototype of a Five-Wheeled Robot and Maneuver Analysis

He Xu, Jinfeng Zhao, Dawei Tan, and Zhenyu Zhang

College of Mechanics and Electrics Engineering, Harbin Engineering University,  
150001 Harbin, China  
railway\_dragon@163.com, zaozongceng@163.com

**Abstract.** An innovative asymmetrical prototype of a five-wheeled robot is first proposed with reconfiguration features. Based on mechanical prototype, the fifth wheel with slippage measurement is then rendered for details. In terms of maneuver features, different control strategy and steering performance for asymmetrical prototype with rhomboid shape have been discussed analytically. At last some tests including reconfiguration tests, straight-line motion and traversing tests have been implemented.

**Keywords:** mobile robot, asymmetrical prototype, maneuver, reconfiguration.

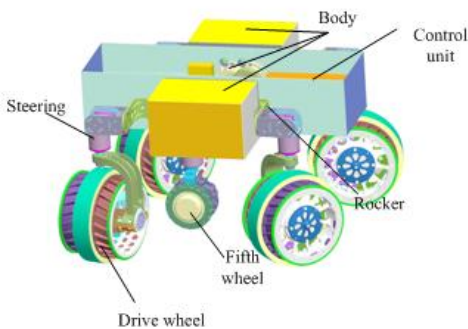
## 1 Introduction

Four-wheeled robot usually has symmetrical shape with two wheels in the front and two wheels in the back, e.g. the Tri-star[1], RATLER[2], SRR rover[3], Micro5[4], Spirit[5] and Lunokhod[6], etc. Fortunately, asymmetrical prototype sometimes indicates unique features. Prof. Shigeo Hirose studied a prototype of four-wheel rover, Rhomboid Rover-I, proposing a rhomboid four wheels vehicle arrangement as a mechanism to manifest terrain adaptability compared to the six wheels vehicle. This mechanism distributes its weight over the four wheels by asymmetrical links and with capability of point-turn without slippage [7]. Dr.Du presented a rover with an integral and statically indeterminate structure, six independent propulsive wheel and asymmetric configuration of wheels at the two sides of rover body. Thus the lunar rover can traverse the ditches on fine-grained surface soil [8]. And an approach to increase or decrease the normal forces of wheels by torsion stiffness and the initial angles of suspension devices is suggested, then suspension can be used to fit the size of fairing [9]. Based on four-wheeled-three-axised structure, a asymmetric rover has been built with only minimum number of four wheels to realize the performance of six-wheeled rover, and its triangle-shaped supporting structure leads to a stable status of a rover. with a minimum number of rounds at home and abroad to achieve the widely used 6 mobile system with three-axis performance, and to leave any one wheel, the other three wheels can form a stable triangular support structure.

An asymmetrical prototype of five wheeled robot is proposed in section 2. The maneuver analysis of asymmetrical prototype of five-wheeled mobile robot is investigated in section 3, and prototype test has been implemented in section 4. And conclusion is drawn in section 5.

## 2 System Design

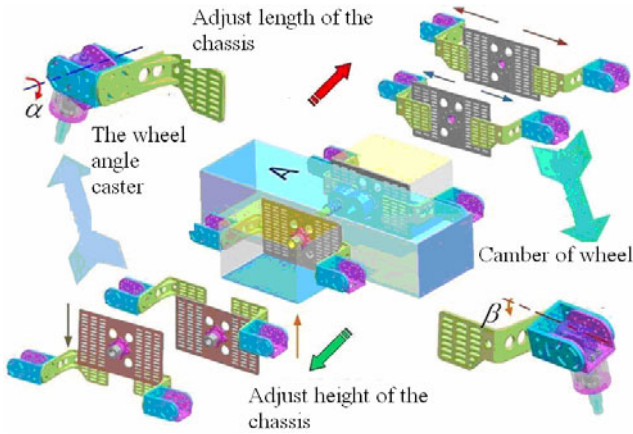
In order to meet the need of the task, conceptual design , program implementation, configuration of multi-objective optimization and interactive integration , whole robot have been devised as been shown in the figure 1.the whole robot is mainly consist of several part, body, control unit, rocker, fifth wheel, steering and drive wheel. Rocker can adjust drive wheel up and down and the steering can make drive wheel rotate because of this design the whole robot is reconstruction. What is more the robot can over-cross the obstacles and climb on slope.



**Fig. 1.** The whole structure of the robot

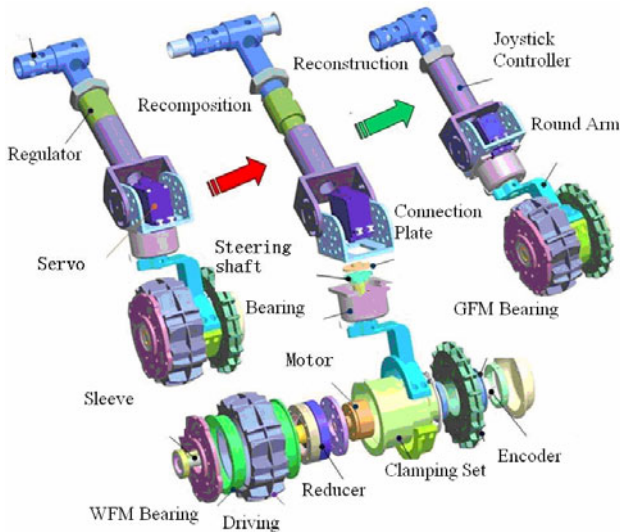
Robot with five wheels adopt suspension system, that is two separate differential input gear shaft are connected to the corresponding rocker and the fifth round of the institutions set can rotate around differential shaft and adapt to changes in terrain. As been see in figure 2, side rocker are consist of these parts the corresponding side of the body, a pair of left and right arm and the regulator component of the angle of inclination before and after. Arm along the side of the body can stand inside to connect the longitudinal and lateral movement, to achieve the chassis adjustment function. each wheel can rotate around two hinged regulator axis of angle regulator component, that is caster and camber, and improve the contact condition of wheel and ground ,and ensure the nature of the wheel's straight .At the same times, unique three-box body fuselage, the main body provide the necessary stable platform for scientific instruments and equipment and on both sides of the body is part of the rocker body not only accommodate the electronic control unit, saving space, but also achieve the functional requirements, improve the dynamic characteristics,. From above we can see that e task-oriented mobile robot can be reconstructed to meet the rugged terrain and various needs.

Driving wheel of the auxiliary support agencies, in the figure 3 (the fifth wheel) consist of the joystick device, steering wheel device and it can rotate around differential drive shaft in the vertical plane to accommodate the topography. Master device from the wheel mainly is composed of two sub-round, with round form inside drive, and are parallel and coaxial installed on both sides of the clamping sleeve.



**Fig. 2.** Reconfigurable robot chassis suspension

Active sub-round achieve to coordinate with other four wheels by the process that the brushless DC motor(EC 32 flat series) connected by harmonic reducer sleeve (XB3-2 series), provide driving torque and than the servo drive (Whistle series) precise control. Sub-wheel driven by the clamping sleeve bearings set in the outer circle on the right side get passive rotation around the topography and make the a fixed encoder (OME-500) move to be relatively accurate measurement of mobile robot the distance traversed. What is more it can further calculate or estimate some of the necessary mobile robot performance parameters, such as the slip rate. From above we see that the fifth wheel mainly have two effects (1) increasing drive of the robot (2)measuring the skip rate of robot and so on



**Fig. 3.** The fifth wheel with slippage measurement

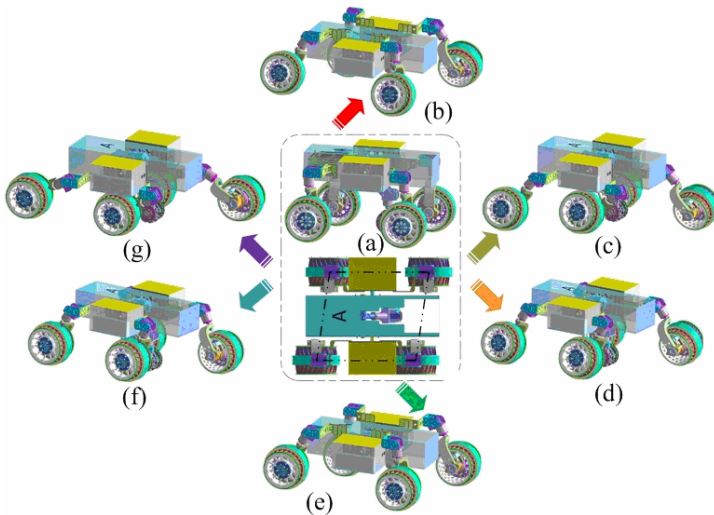
Some important structure parameters are listed in the table 1. Structural design of the robot's body is essentially an integrated mechatronic system design process. In full consideration of the overall process control system, the specific mechanical structure design was made. Differential structure is the important part for adjusting the balance. Reconstruction is the biggest character of this robot.

**Table 1.** The important Parameter of robot

Parameter	data	Parameter	data
Effective length of the chassis /m	0.323~0.399	Fifth wheel height /m	0.0306
Chassis height adjustment /m	-0.0483~0.069	Equivalent axilbase /m	0.3164~0.7764
Front wheel caster angle /deg	-90~1	Equivalent wheelbase /m	0.2971~0.4466
Front wheel camber angle /deg	-18~5	Height /m	0.0383~0.2753
back wheel caster angle /deg	-1~90	longitudinal stability /deg	39.995~77.21
back wheel camber angle /deg	-18~5	Longitudinal stability of the robot /deg	3.662~32.24

### 3 The Analysis of Prototype of Asymmetrical Prototype

Because the robot have the character of reconstruction lots of configurations are acquired .the diamond is the special shape so this configuration was devised that the whole robot's projection plane is diamond.This configuration was made by many ways. In the figure 4 six different patterns are displayed.



**Fig. 4.** Different configurations

This article is the asymmetric mobility of the robot. Robot has multiple motor modes, such as omni-direction, skip mode, turn point, brake mode and difference mode. This chapter mainly introduces the relationship with different wheel speed in all kinds of typical pattern in the asymmetry of wheels turning. The distance from throughout the robot's speed Instantaneous center  $I$  to rotation center  $R$  is  $r$ . Through the  $r$  we can get different wheels' distance to speed Instantaneous center  $I$ . finally we can get the relationship with different wheel speed.

### 3.1 Ackerman Mode

Shown in figure 5,  $\delta_1, \delta_4$  denote the wheel 1 and wheel 2 Corners.  $R$  is the reference point and  $r$  is the distance between reference point and speed Instantaneous center. According to the  $\delta_1, \delta_4$ , we can get the location of the speed Instantaneous center  $I$  and get the Direction of wheel 2, wheel 3, fifth wheel. and reference point  $R, \delta_2, \delta_3, \delta_5$  and  $\delta$ . Now we will get the relationship with different wheels and reference radius  $r$ .  $r_1, r_2, r_3, r_4$  and  $r_5$  denote the radius of wheel 1, wheel 2, wheel 3, wheel 4 and fifth wheel to speed Instantaneous center  $I$ .

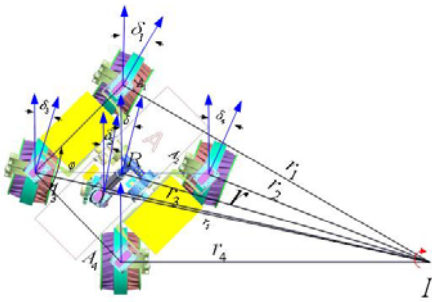


Fig. 5. Ackerman mode

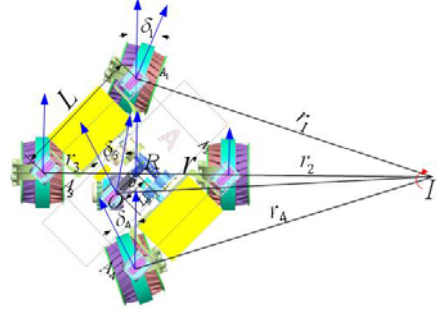


Fig. 6. Double Ackerman mode

$$\therefore \angle RA_4I = 0.5\pi \tag{1}$$

$$r_4 = r_4 = \sqrt{r^2 - L^2 \sin^2(0.5\varphi)} \tag{2}$$

$L$  denote the distance of  $A_1 A_3$  and  $\varphi$  denote the Angle of the  $A_1 A_3$  and  $A_3 A_4$

$$\delta = \arccos \frac{r_4}{r} \tag{3}$$

$$r_1 = \sqrt{r_4^2 + 4L^2 \sin^2(0.5\varphi)} = \sqrt{r^2 + 3L^2 \sin^2(0.5\varphi)} \tag{4}$$

$$\delta_1 = \arccos \frac{r_4}{r_1} \tag{5}$$



Shown in figure 5

$$r_3 = \sqrt{L^2 + r^2 - L^2 \sin^2(0.5\varphi) + 2L \cos(0.5\varphi) \sqrt{r^2 - L^2 \sin^2(0.5\varphi)}} \tag{6}$$

According to the sine Theorem

$$L / \sin \delta_3 = r_3 / \sin(\pi - 0.5\varphi) \tag{7}$$

$$\delta_3 = \arcsin \frac{L \sin(0.5\varphi)}{r_3} \tag{8}$$

The same with above

$$r_2 = \sqrt{L^2 + r^2 - L^2 \sin^2(0.5\varphi) - 2L \cos(0.5\varphi) \sqrt{r^2 - L^2 \sin^2(0.5\varphi)}} \tag{9}$$

$$\delta_2 = \arccos \frac{L \sin(0.5\varphi)}{r_2} \tag{10}$$

$$r_5 = \sqrt{r^2 + L_5 - 2rL_5 \cos(\arccos \frac{r_4}{r} - \theta)} \tag{11}$$

$\theta$  denote the angle of OR and R  $A_4$  and  $L_5$  denote the distance of OR

$$L_1 = \sqrt{L_5^2 + L^2 \sin^2(0.5\varphi) - 2L_5 L \sin(0.5\varphi) \cos \theta} \tag{12}$$

$L_1$  denote the distance from O to  $A_4$

The same with above

$$\delta_5 = \arccos \frac{r_5^2 + r_4^2 - L_1^2}{2r_5 r_4} \tag{13}$$

### 3.2 Double Ackerman Mode

Arranging the speed Instantaneous center  $I$  on the line of the central axis, R is the reference point and  $r$  is the distance of the reference point to the speed Instantaneous center  $I$ .  $r_1, r_2, r_3, r_4$  and  $r_5$  denote the distance of wheel 1, wheel 2, wheel 3, wheel 4 and fifth wheel to the speed Instantaneous center  $I$ .  $\delta_1, \delta_2, \delta_3, \delta_4, \delta_5$  and  $\delta$  denote the Corners of the wheel 1, wheel 2, wheel 3, wheel 4, fifth wheel and reference point, as shown in the figure 6.

$$\delta_2 = \delta_3 = \delta = 0, \delta_1 = -\delta_4 \tag{14}$$

$$r_2 = r - L \cos(0.5\varphi) \tag{15}$$

$$r_3 = r + L \cos(0.5\varphi) \tag{16}$$

L denote the distance of  $A_1 A_3$  and  $\varphi$  denote the Angle of the  $A_1 A_3$  and  $A_3 A_4$

$$r_1 = \sqrt{L^2 + (r - L \cos(0.5\varphi))^2 + 2(r + L \cos(0.5\varphi))L \cos(0.5\varphi)} \tag{17}$$

$$\delta_1 = \arctan \frac{L \sin(0.5\varphi)}{r} \tag{18}$$

The same with the above

$$r_4 = \sqrt{r^2 - L^2 \sin^2(0.5\varphi)} \tag{19}$$

$$r_5 = \sqrt{r^2 + L_5^2 + 2rL \cos \theta} \tag{20}$$

According to the sine Theorem

$$L_5 / \sin \delta_5 = r / \sin \theta \tag{21}$$

$$\delta_5 = \arcsin(L_5 \sin \theta / r) \tag{22}$$

### 3.3 Point-Turn Mode

Shown in the figure 7, This mode must let the whole robot rotate around the middle point that is R so we known that the reference point R and speed Instantaneous center at the same point.

From the figure 7, we can get that wheel 1 and wheel 4 have the same speed but the direction is opposite. The wheel 2 and the wheel 3 have the same condition with the wheel 1 and wheel 4. It is easy for us to get the relationship as follow:

$$\delta_1 = 0.5\pi, \quad \delta_4 = -0.5\pi, \quad \delta_3 = \delta_4 \tag{23}$$

$$r_1 = r_4 = L \sin \varphi \quad r_3 = r_2 = L \cos \varphi \quad r_5 = L_5 \tag{24}$$

L denote the distance of  $A_1 A_3$  and  $\varphi$  denote the Angle of the  $A_1 A_3$  and  $A_3 A_4$

$\theta$  denote the angle of OR and R  $A_4$  and  $L_5$  denote the distance of OR

$$\delta_5 = 0.5\pi - \theta \tag{25}$$

### 3.4 Omni-Direction

This mode in figure 8, we arrange that all wheels have the same Corner  $\delta$ , so we can get speed Instantaneous center at the Infinity. omni-direction is the best mode to reduce the energy and it is a easy mode

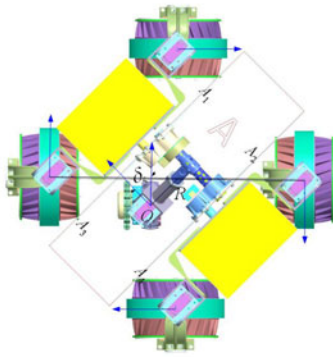


Fig. 7. Turn point mode

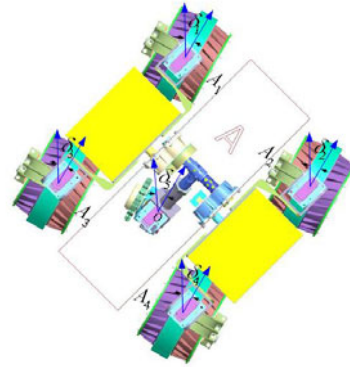


Fig. 8. Omni-direction

### 3.5 Skip Mode

Shown in the figure 9, speeds of wheel 1 and wheel 4 are zero and wheel 2, wheel 3 get the opposite and the equal speed so we know that reference point and speed Instantaneous center at the same point. We can get the speed direction of the fifth wheel  $\delta_5$ . From the figure 9, the relationship with all wheel's speed was got as follows:

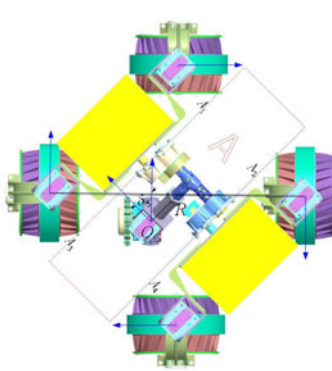


Fig. 9. Skip mode

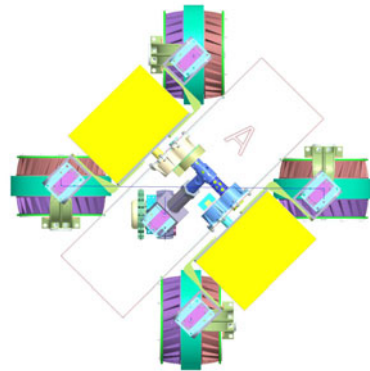


Fig. 10. Brake mode

$$r_3 = r_2 = L \cos(0.5\pi) \tag{26}$$

$$r_5 = L_5 \tag{27}$$

$$\delta_5 = 0.5 - \theta \tag{28}$$

Where  $L$  denote the distance of  $A_1$ ,  $A_3$  and  $\varphi$  denote the Angle of  $A_1A_3$  and  $A_3A_4$   $\theta$  denote the angle of  $OR$  and  $R A_4$  and  $L_5$  denote the distance of  $OR$

### 3.6 Brake Mode

Shown in the figure 10, wheel 1 and wheel 4 have the same speed but the direction is opposite the same with it wheel 2 and wheel 4 have the same condition .From this condition ,the robot can not move so we call it brake mode

### 3.7 Difference Mode

This mode we arrange that wheel 1, wheel 4, have the zero speed and the wheel 2, wheel 3 have the speed but the speed of wheel 3 is two times than the wheel 2. From the figure 11, speed Instantaneous center is identified at the line of the central axis. Because the speed of wheel 3 is two times compared with wheel 2, the location of the speed Instantaneous center can be got. According to the speed Instantaneous center, the speed direction of fifth wheel is made. we can get the equation as follows:

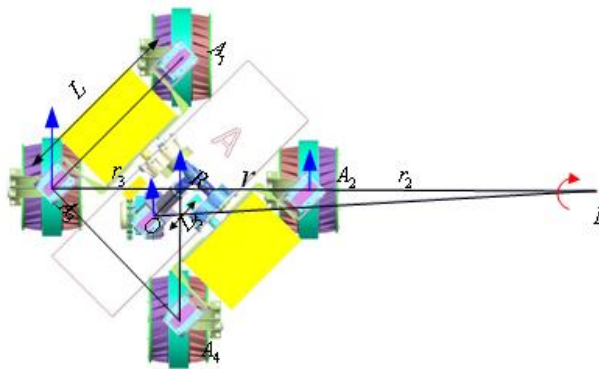


Fig. 11. Difference mode

$$r = 3L \cos(0.5\varphi) \tag{29}$$

$$r_3 = 4L \cos(0.5\varphi) \tag{30}$$

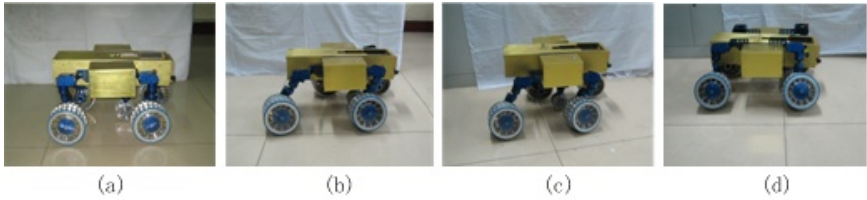
$$r_2 = 2L \cos(0.5\varphi) \tag{31}$$

$$r_5 = \sqrt{r^2 + L_5^2 + 2L_5 r \cos \theta} \tag{32}$$

$$\delta_5 = \theta \tag{33}$$

### Prototype Test

Reconfiguration prototypes of asymmetrical prototype of mobile robot are pictured in figure 12, and the relative data are shown in table 2.



**Fig. 12.** Asymmetrical prototypes

**Table 2.** Class diamond configuration of the theoretical and actual value

Parameter	the chassis length (a)		Comprehensive regulation (b)		Comprehensive regulation (c)	
	theory	Practice	theory	Practice	theory	Practice
$l_{14}$	0.1995	0.1995	0.1995	0.1995	0.1995	0.1995
$l_{23}$	0.1615	0.1615	0.1615	0.1615	0.1615	0.1615
$h$	0	0	0	0	0.0483	0.04873
$\alpha_l$	0	0	-30	-30	-30	-30
$\beta_l$	0	0	-14	-14	-8	-8
$\alpha_r$	0	0	30	30	30	30
$\beta_r$	0	0	14	14	8	8
$l_b$	0.150	0.150	0.135	0.135	0.170	0.170
$H_C$	0.2270	0.2155	0.1969	0.1860	0.2484	0.2371
		0.2150		0.1858		0.2365
		0.2153		0.1855		0.2361
$H_C$	5.17%		5.65%		4.76%	



**Fig. 14.** Diamond-shaped configuration of five straight wheel class performance test ( $\alpha_l = -18^\circ$  and  $\alpha_b = 18^\circ$ )



**Fig. 15.** Five diamond wheel type configuration through the rough ground test ( $\alpha_l = -18^\circ$  and  $\alpha_b = 18^\circ$ )

## 5 Conclusions

The prototypes of various configurations with asymmetrical configuration and rhomboid-like configuration are compared in terms of maneuver action theoretically. Lastly we do lots of experiment for robot including the prototype adaptive capacity, walking ability through nature climbing and mobility patterns. The results show that prototype can be reasonable and practicable to re-form a variety of configurations and better ground adaptability (discrete differential device function), the more impaired climbing ability. The robot can complete seven mobility patterns and this device can meet the demand.

## Acknowledgments

This work is supported by National Science Foundation of China under Grant 60775060 and 6067406, the Foundation Research Fund of Harbin Engineering University under Grant HEUFT07027, the Research Fund for Doctoral Program of Higher Education of China under Grant 20070420870 and the Academy of Finland under Grant 214144.

## References

1. <http://www.masaakix.interlink.or.jp/apollo/apollo-15/ap15-lrv.html>
2. <http://www.cs.cmu.edu/~lri/nav96.Shtml>
3. Huntsberger, T., Baumgartner, E., Aghazarian, H., Cheng, Y., Schenker, P., Leger, P., Iagnemma, K., Dubowsky, S.: Sensor fused autonomous guidance of a mobile robot and applications to Mars sample return operations. In: Proceedings of the SPIE Symposium on Sensor Fusion and Decentralized Control in Robotic Systems II, vol. 3839, pp. 2–8 (1999)
4. Kubota, T., Kuroda, Y., Kunii, Y., et al.: Micro planetary rover“Micro5”. In: 5th International Symposium on Artificial Intelligence, Robotics and Automation in Space (ISAIRAS 1999), June 01-03, pp. 373–378 (1999)
5. Volpe, R.: Rover Technology Development and Mission Infusion. In: IEEE ICRA 2005 Planetary Rover Workshop, Barcelona, Spain (April 2005)
6. <http://www.planetary.org/rrgtm/tpr-rover-rus.html>
7. Hirose, S.: Fundamental Considerations for the Design of a Planetary Rover. In: IEEE International Conference on Robotics and Automation, pp. 1939–1944. IEEE, Tokyo, 0-7803-1965-6/95 \$4.00 01995
8. Du, J.-J., Liu, D., Yuan, L.-H., Zhao, Z.: Research on kinematics and dynamics of lunar rover with new configuration. Journal of Harbin Institute of Technology(English) 12(suppl. 2), 131–135 (2005)
9. Dun, L., Jian-jun, D., Zhiping, Z., et al.: A new configuration of mobile system of lunar rover and its drive control study. In: International Lunar Conference 2005, Canada (2005)

# Development of Controllable Two-Way Self-locking Mechanism for Micro In-Pipe Robot

Xue Yong\*, Shang Jian-zhong, Luo Zi-rong, Qiao Jin-wei, and Cheng Chen

Department of Mechatronics Engineering, National University of Defense Technology  
Changsha 410073, China  
Tel.: 0731-84535584  
xueyong\_11@163.com

**Abstract.** A controllable two-way self-locking mechanism was developed in this paper to improve the tractive ability of inchworm micro In-pipe robot and reverse its moving direction. Firstly, the robot structure was proposed and its relevant locomotive functions were introduced. Then the mechanism was analyzed and its mathematical model was built to deduce the conditions of direction reversing and diameter variation adapting. Finally a robot prototype was developed based on the results derived above, and its related experiments were carried to test these conditions.

**Keywords:** Micro In-pipe Robot; Two-way Self-locking Mechanism; Diameter Variation Adapting.

## 1 Introduction

As an efficient inspection method for micro pipes, in-pipe micro robots have been a research hotspot home and abroad [1]. By now, they can be generally divided into wheel type, inchworm type and vibration type. Wheel type contains straight-line type [2,3] and screw type. Iwao Hayashi and his colleges in Tokyo Institute of Technology were the first ones to develop screw type in-pipe micro robot [4,5]. The magnitude of the pull of wheel type robot depends on the static friction, and the static friction rises with the increasing pressure against the wall. Therefore, the pull cannot go up as will because of the finitude of pressure. Inchworm type robot attracts most attention in this field [6]. It mainly composes of two crutches at each end and driving mechanism in the middle. The wheel type and inchworm type robots get power and control signals through cables, requiring a huge tractive force. A vibration type micro In-pipe robot based on wireless energy and signal transmission was developed.

This paper analyzes micro In-pipe robot with the background of auto-inspection for heat transfer tubes of nuclear reactor boilers in nuclear power stations and nuclear submarines, mainly aiming at huge driving forces, two-way movements and fast long distance motion.

Based on cam's self-lock mechanism, a new controllable two-way self-locking structure has been designed, enabling micro In-pipe robot realize two-way movement

---

\* Corresponding author.

and own a huge tractive force, simultaneously self-adjust suit the diameters variation to a certain degree. This paper has expatiated and analyzed the direction reversing and self-locking mechanisms, discussing conditions each parameters need to meet after building a mathematical model. Then an example is afforded, and a robot prototype is accomplished.

## 2 Telescopic Micro In-Pipe Robot Based on Controllable Self-locking Mechanism

This robot is designed on cam's self-locking mechanism, and the self-locking structures at each end enable two-way locking, thus the robot can move forward and backward by changing the locking direction. Fig.1 shows a gait cycle containing two phases of the robot. Fig.1(a) expresses a gait period of motion for forward travel, and Fig.1(b) expresses a gait period for backward travel.

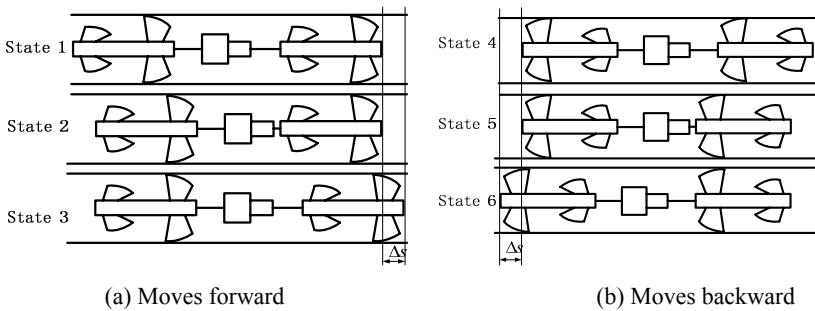


Fig. 1. Locomotion of the robot

When it moves forward, the front cams are stretching while the back cams are contracted in each self-locking structure, enabling lock the backward direction. When the telescopic mechanism contracts, the front self-locking structure gets locked while the back one is pulled forward from state 1 to state 2. When the telescopic mechanism stretches out, the back self-locking structure gets locked while the front one is pulled forward from state 2 to state 3. In this forward gait cycle, the robot's displacement is  $\Delta s$ . When the robot moves backward, the front cams are contracted while the back cams are stretching in each self-locking structure, enabling the robot lock the forward direction. The motion of telescopic mechanism is same argument.

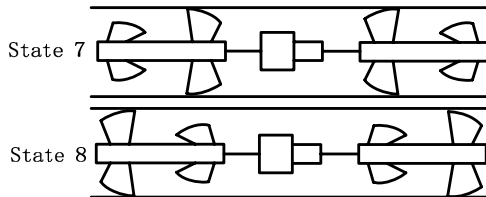


Fig. 2. Localization of the robot



When the robot is in any state shown in Fig.2, localization is gained. What shows in state 7 and state 8 is one locking structure lock left while the other one lock right.

### 3 Controllable Two-Way Self-locking Mechanism

#### 3.1 Cam's Unilateral Self-locking Principle

As the base of unilateral self-locking design, cam's unilateral self-locking principle is shown in Fig.3. Point  $O$  is the rotating center of the cam, and point  $P$  is the contact point between cam and pipe. If the cam tends to move left shown in Fig.3(a), the cam would rotate clockwise by the friction  $f$  of pipe wall. Meanwhile, the radius  $\rho$  of the cam, and the pressure  $N$  between the pipe wall and the cam, as well as the friction all increase. Thus the cam is locked, restricting moving left. If the cam tends to move right shown in Fig.3(b), the cam would rotate anti-clockwisely by the friction  $f$  of pipe wall. Meanwhile, the radius  $\rho$  of the cam, and the pressure  $N$  between the pipe wall and the cam, as well as the friction all decrease. Thus the cam is unlocked, and can move right.

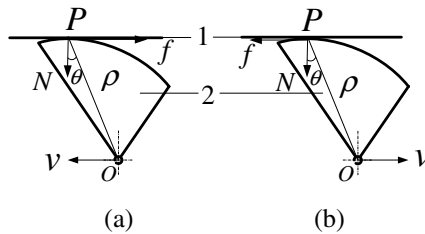


Fig. 3. Cam's Self-locking Principle

From the above we can get that if the arc of the cam meets self-locking conditions, the cam can lock itself. If the friction coefficient between the cam's arc and the pipe wall is assumed  $\mu$  and the friction is assumed  $f$ , then we can get the self-locking condition.

$$N \tan \theta < f = N\mu . \tag{1}$$

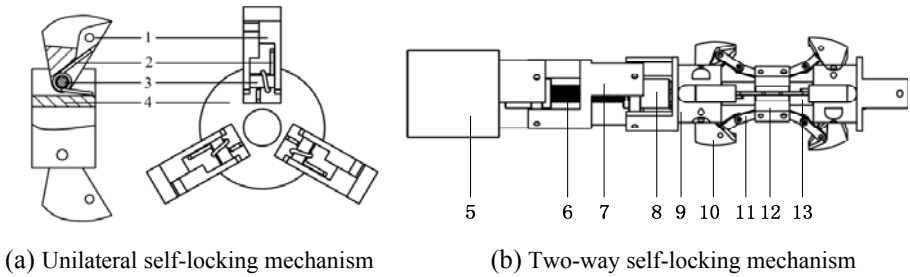
It can be further written as

$$\tan \theta < \mu . \tag{2}$$

#### 3.2 Structural Design

The self-locking mechanism is depicted in Fig.4. And Fig. 4(a) shows the unilateral self-locking structure. The cam is located at the motor-mount circumferentially  $120^\circ$  apart from each other and connects together by the pin roll. The torsional spring is installed in the pin roll, forcing the cam on the pipe wall all the time. Fig.4(b) shows two-way self-locking mechanism. Two unilateral self-locking mechanisms are tied by a lead screw, the moving nut and the cam is tied by connecting bars, the motor-mount and unilateral self-locking mechanism are tied by a gimbal joint, and the motor shaft and the lead screw are tied by a flexible shaft and clutch. If motor and unilateral self-locking

mechanism were connected directly, the single unit of the In-pipe robot would be long, increasing the turning radius. The gimbal joint and flexible shaft meet motor's driving requirement and shorten single unit's length and the turning radius.



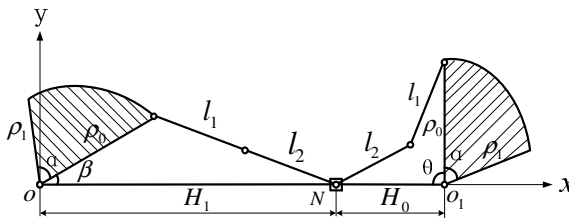
**Fig. 4.** Self-locking mechanism. 1: cam; 2: torsional spring; 3: pin roll; 4: fix mount; 5: motor and its mount; 6: flexible shaft; 7: gimbal shaft; 8: clutch; 9: fix mount; 10: cam; 11: connecting bar; 12: moving nut; 13: lead screw.

This two-way self-locking mechanism's principle of work is as follows: motor drives the lead screw through the flexible shaft, and the lead screw drives the moving nut longitudinally. If the self-locking mechanism locks left, the moving nut travels right to the low order end, stretching the left connecting bar to contract the left cam, while the right connecting bar is released and the right cam is stretched out by the torsional spring and presses on the pipe wall, locking left. If the self-locking mechanism locks right, the moving nut travels left to the low order end, stretching the right connecting bar to contract the right cam, while the left connecting bar is released and the left cam is stretched out by the torsional spring and presses on the pipe wall, locking left.

## 4 Analysis of Two-Way Controllable Self-locking Mechanism

### 4.1 Analysis of Direction Reversing Condition

The mathematical model of two-way self-locking mechanism is depicted in Fig.5. To realize direction reversed, self-locking mechanism should meet conditions as follows: Only one set of cams can work. That is to say, when one set has touched the inner pipe wall the other one should be drawn back to avoid contact with the pipe wall. As illustrated in Fig.5, the right set is working and the left one has been drawn back.



**Fig. 5.** Analysis on direction reversed condition

To control and compute conveniently, the moving nut is only in two positions. One is that makes the back cam is entirely contracted and the front cam is entirely stretching out when the robot moves forward, another is that makes the front cam is entirely contracted and the back cam is entirely stretching out when the robot moves backward. To achieve this goal,  $H_1$ 、 $H_0$  need to satisfy

$$\begin{cases} H_1 \geq \rho_0 \cos(90^\circ - \alpha) + \sqrt{(l_1 + l_2)^2 - \rho_0^2 \sin^2(90^\circ - \alpha)} \\ H_0 \leq \sqrt{(l_1 + l_2)^2 - \rho_0^2} \end{cases} \quad (3)$$

### 4.2 Analyses of the Singular Positions

Diameters of practical pipes are variable and even in the same pipe because external force leads to change the diameter at some place, thus In-pipe robots are required to fit this situation. The In-pipe robot designed in this paper can suit the scale of diameter adaptation shown in Fig. 6.

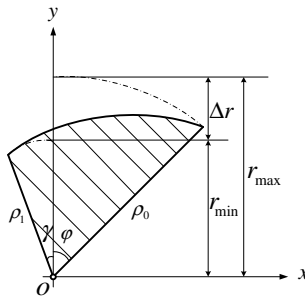


Fig. 6. Analyses on the scale of diameter adaptation

If the cam turn clockwise and  $\gamma=0^\circ$ , the minimum radius the robot can suit is  $r_{\min} = \rho_1$ . If the cam turn anti-clockwise and  $\varphi=0^\circ$ , the maximum radius the robot can suit is  $r_{\max} = \rho_0$ . So the variation scale of diameter the robot can suit is  $2\Delta r$ .

According to the analysis above, in-pipe robot suits different diameters by rotating cams. To get a maximum changing scale, cams are required to rotate without disturbance, so it is quite important to analyze the singular positions.

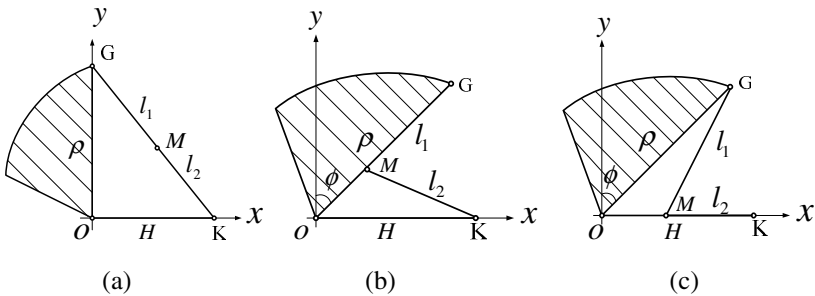


Fig. 7. Analyses on the singular positions

Three possible singular positions are shown in Fig.7. Fig.7(a) shows two connecting bars are in a line, i.e. Point  $G$ , Point  $M$  and Point  $K$  are all in a line, forming a singular position of four bar linkage mechanism, thus the cam can't rotate. Fig. 7(b) shows a singular position where Point  $M$  is in line  $OG$  when the cam rotates  $\phi$  ( $\phi < \alpha$ ), because the point  $M$  could not cross the cam in practice. And Fig. 7(c) shows a singular position where Point  $M$  is in line  $OK$  when the cam rotates  $\phi$  ( $\phi < \alpha$ ), because the point  $M$  could not cross the lead screw in practice. The two singular positions in Fig. 7(b) and (c) would not influence the cam's stretching if  $\phi > \alpha$ , where  $\alpha$  is the cam's expanding angle.

### 4.3 Connecting Bars Design

The mathematical model of connecting bars is shown in Fig. 8.  $Oxy$  is the fixed rectangular coordinate system. And  $\alpha$  is the cam's expanding angle. Points  $O$ ,  $G$ ,  $M$  and  $K$  are the four hinge points.  $\theta$  is the angle between line  $OG$  and  $y$  axis, and  $\phi$  is the angle between line  $OM$  and  $y$  axis.

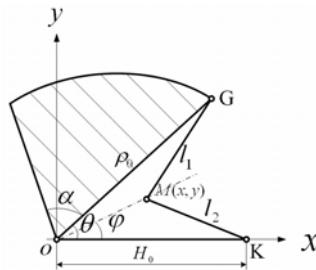


Fig. 8. Mathematical model of connecting bars

As described above, in order to suit maximum diameter, i.e.  $\theta = 90^\circ$  and avoiding the singular position shown in Fig. 7(a), parameters related need to meet

$$H_0 < \sqrt{(l_1 + l_2)^2 - \rho_0^2} \quad (H_0 > 0) . \tag{4}$$

If the diameter of pipe changes, the cam would rotate to suit it. The constraint conditions of  $M$  can be described as

$$\begin{cases} (x - \rho_0 \cos \theta)^2 + (y - \rho_0 \sin \theta)^2 - l_1^2 = 0 \\ (x - H_0)^2 + y^2 - l_2^2 = 0 \end{cases} . \tag{5}$$

Then, the Eq. (5) can be written as

$$At^2 + Bt + C = 0 , \tag{6}$$

where

$$\begin{cases} t = \frac{y}{x} \\ A = (l_1^2 - l_2^2 + H_0^2 - \rho^2)^2 - 4(l_2^2 - H_0^2)\rho^2 \sin^2 \theta \\ B = [4H_0(l_1^2 + l_2^2 - H_0^2 - \rho^2) - 8\rho(l_2^2 - H_0^2)\cos\theta]\rho \sin \theta \\ C = 4(\rho \cos \theta - H_0)(H_0 l_1^2 - H_0 \rho^2 - l_2^2 \rho \cos \theta + H_0^2 \rho \cos \theta) \\ \quad + (l_1^2 - l_2^2 + H_0^2 - \rho^2)^2 \end{cases} \quad (7)$$

From Eq. (6) we get

$$t = f(l_1, l_2, H_0, \theta) \quad (8)$$

When  $\theta = 90^\circ$ , the robot can reach maximum diameter. When  $\theta = 90^\circ - \alpha$ , it will fit minimum one. Thus  $\theta \in [90^\circ - \alpha, 90^\circ]$  is assumed. If  $\theta$  belongs to it, to forbid the singular positions in Fig. 7(b) and (c),  $\varphi$  will be satisfied with

$$0 \leq \tan(\varphi) \leq \tan(\theta) \quad (9)$$

Then Eqn.(9) can be written as Eqn.(10) in terms of the locomotion consistency of four bars linkage mechanism.

$$0 \leq \tan(\varphi) \leq \tan(90^\circ - \alpha) \quad (10)$$

Substituting  $\theta = 90^\circ - \alpha$  into Eq.(7), from Eq.(7) and Eq.(8), we get

$$\begin{cases} \varphi = \arctan\left(\frac{y}{x}\right) = \arctan(t) \\ t = f(l_1, l_2, H_0) \end{cases} \quad (11)$$

However, the solution of  $t = f(l_1, l_2, H_0)$  may be imaginary if  $l_1$ ,  $l_2$ ,  $H_0$  are endowed unreasonable values when  $\theta = 90^\circ - \alpha$  is restricted. To avoid this,  $B^2 - 4AC \geq 0$  is firstly satisfied, i.e.

$$(l_1 - l_2)^2 \leq (\rho_0 \sin \alpha - H_0)^2 + \rho_0^2 \sin^2 \theta \quad (12)$$

From the above, to suit the maximum changing scale of diameters,  $l_1$ ,  $l_2$ ,  $H_0$  should be satisfied with

$$\begin{cases} 0 < H_0 < \sqrt{(l_1 + l_2)^2 - \rho_0^2} \\ (l_1 - l_2)^2 \leq (\rho_0 \sin \alpha - H_0)^2 + \rho_0^2 \sin^2 \theta \\ 0 \leq \arctan(f(l_1, l_2, H_0)) \leq 90^\circ - \alpha \end{cases} \quad (13)$$

## 5 Design Calculation and Experiments

### 5.1 Design Calculation

The controllable two-way self-locking mechanism is designed based on diameters from  $\Phi 15\text{mm}$  to  $\Phi 20\text{mm}$ . From the theoretical analysis about the variation scale the robot can suit, we set  $\rho_0=10\text{mm}$ ,  $\rho_i=7.5\text{mm}$ , and  $\alpha=50$  ( $\theta=40^\circ$ ) after computing cam's self-locking condition. To meet assembly dimension  $H_0=8\text{mm}$  is set, and from Eq.(6),  $l_1+l_2>12.8\text{mm}$ , thus the round number is 13mm. Substituting values of  $H_0$ ,  $\rho_0$ ,  $\theta$  into Eq.(9) and Eq.(10), all values of  $\varphi$  are calculated in the environment of MATLAB when  $l_1$  changes from 1mm to 12mm and shown in Fig.9, where  $t_1$ ,  $t_2$  are two solutions of Eq.(9). To meet  $0^\circ \leq \varphi \leq 40^\circ$ ,  $l_1=7.5\text{mm}$  is set from Fig.9, then  $l_2=5.5\text{mm}$ .

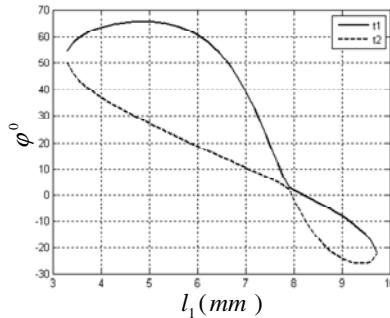


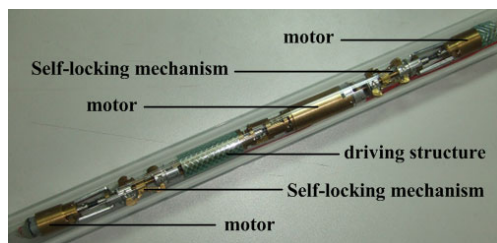
Fig. 9. Curve of  $\varphi$

### 5.2 Experiments Demonstration

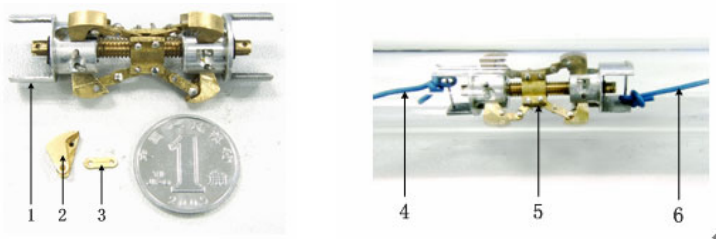
A robot prototype was accomplished, shown in Fig. 10. The controllable two-way self-locking mechanism was designed with diameters from 15mm to 20mm, shown in Fig. 11(a). Tractive force and direction-reverse experiments are depicted in Fig. 11(b). The pull was offered by the rope. First of all, the robot was put into a pipe with a nominal diameter of 18mm while the lead screw was regulated to lock leftward. A pull to the right was acted on the right part of the mechanism, thus it can move rightward. Then a pull to the left was acted on the left part. It locked itself. Then a tension pick-up YHZL-91006 was installed in the rope. The pull was increased step by step. Before it arrived at the maximum value 2.5kg, the mechanism hasn't slipped obviously. The similar experiments for right locking were carried and some identical results were obtained.

Pipes of different diameters were utilized in experiments, demonstrating that the self-lock mechanism can adjust diameters from 15mm to 20mm.

Experiments demonstrated that the controllable two-way self-locking mechanism can achieve two-way self-locking and gained a relative high tractive force and flexible pipe diameter variation adapting ability.



**Fig. 10.** Robot prototype



(a) two-way self-locking mechanism

(b) experiment

**Fig. 11.** Photos of the robot prototype. 1: Two-way self-locking mechanism; 2: cam; 3: connecting bar; 4: string; 5: pipe; 6: string.

## 6 Conclusions

(1) A two-way self-locking mechanism has been designed, which can change its locking direction and suit the diameters variation to some extent.

(2) The mathematical model was built and its basic conditions were deduced.

(3) A robot prototype was accomplished and its relevant experiments were carried out.

## Acknowledgement

We would like to express our appreciation to the Hi-Tech Research and Development Program (863) of China (2007AA04Z256).

## References

1. Yongshun, Z.: Development of micro In-pipe robot abroad. *J. Robot* (6), 506–507 (2000)
2. Fukuda, T., Hoaokai, H.: Study on Inspection Mobile Robot for the Inside of Pipeline. *J. Transactions of the Japan Society of Mechanical Engineers* 52(477), 1584–1588 (1984)
3. Haiyan, T., Long, X., Zhangjun, S.: Design of running system for straight-line type micro In-pipe robot. *J. Journal of Beijing Institute of Petro-Chemical* 14(2), 33–36 (2006)

4. Hayashi, Iwatsuki, N., Iwashina, S.: The Running Characteristics of a Screw-Principle Microrobot in a Small Bent Pipe. C. In: IEEE Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, pp. 225–228 (1995)
5. Yu, G.: Design and Analysis of screw type In-pipe robot. D. National University of Defense Technology, Changsha (2006)
6. Qiang, L., Lishu, W., Shuhua, Z., Hong, C.: Development of In-pipe robot in the oil gas storage and transportation industry. J. Technology and Equipment of pipe 6, 24–26 (2006)



# Trench-Crossing Capability Analysis of a Reconfigurable Tracked Mobile Robot

Hei Mo<sup>\*</sup>, Shang Jianzhong, Luo Zirong, and Wang Zhuo

Department of Mechatronics Engineering, National University of Defense Technology,  
Changsha 410073, China  
Tel.: 0731-84535584  
heimo\_pla@163.com

**Abstract.** A reconfigurable tracked mobile robot (RTMBot) is presented in this paper based on a parallel four-bar mechanism. Firstly the structure of the robot is introduced. After that a centroid configuration model is proposed to analyze the trench-crossing capability of the robot and plan the locomotion configurations for crossing trenches. Then a virtual robot prototype has been constructed to analyze and simulate the performance of trench-crossing. Finally, a prototype of the reconfigurable tracked mobile robot has been produced, and experiments have been performed to verify the design concept and its embodiment. The results show that RTMBot can reconfigure itself to cross trenches.

**Keywords:** Tracked mobile robot, Reconfiguration, Trench-crossing, Centroid.

## 1 Introduction

As mobile platforms, Tracked mobile robots armed with diversified equipments can enter dangerous or unknown environments to carry out different missions. Usually, tracked mobile robots have excellent obstacle overcoming performance and powerful mobility. They are widely used in the military and security fields, especially in tasks, such as rescue, anti- terrorism, anti-riot, reconnaissance and street-battles[1,2].

The capability of trench-crossing is a main criterion of obstacle overcoming performance of tracked mobile robots. The robot's centroid plays a key role in the process of trench-crossing. Once the centroid gets across the trench, the robot can successfully to do so[3,4]. Currently, the locomotion mechanism of tracked mobile robots includes: 2-parallel-track[5], 4-track[6,7,8] and 6-track[9], etc. Most of them cannot adjust their centroids, which seriously hamper their performance of trench-crossing.

To improve the performance of those tracked mobile robots mentioned above, the IRobot company (USA) developed the iRobot PackBot[10] which can be equipped with one or two pairs of fin tracks to overcome obstacles as shown in Fig.1. Inspired by the tracked structure of PackBot and modular design concept, some other related researches have been carried out[9,11,12].

---

\* Corresponding author.

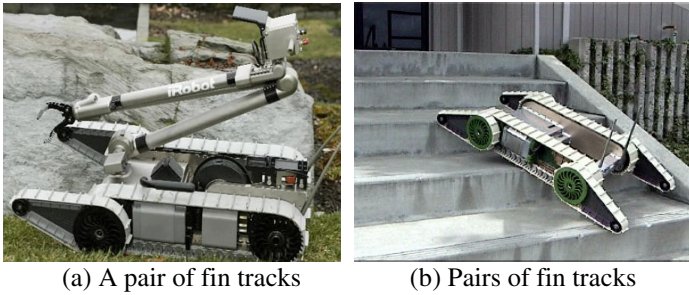


Fig. 1. PackBot

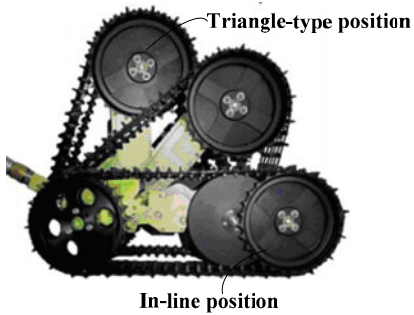


Fig. 2. VGTV



Fig. 3. VIPER

The Inuktun Company in Canada developed a scout tracked mobile robot with changeable track configuration, named VGTV as shown in Fig.2. Usually, it works as a 2-parallel-track tracked mobile robot when tracks both in line. When conferring obstacles, the robot can transform its tracks into a tri-angle type[13].

Fig.3 shows a reconfigurable robot, named VIPER, which developed in Elbit Company of Israel in 2007. With two working modes, wheeled and tracked, the robot integrated the virtue of wheeled and tracked mobile robot by a track & wheel transformation mechanism[14].

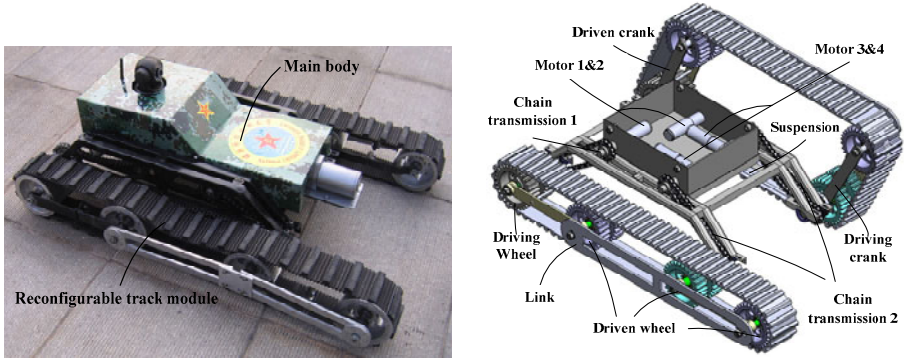
The reconfigurable tracked mobile robots, such as Packbot, VGTV and VIPER can change their configurations and shapes to improve the capability of trench-crossing, but their mechanisms are complex, which increase the difficulties in maintenance and control. A reconfigurable tracked mobile robot (RTMBot) with simple mechanism is presented in this paper based on a parallel four-bar mechanism. It can change its own centroid with reconfigurable track modules, which effectively increase the capability of trench-crossing.

## 2 Structure and Centroid Configuration Model

### 2.1 Structure Design

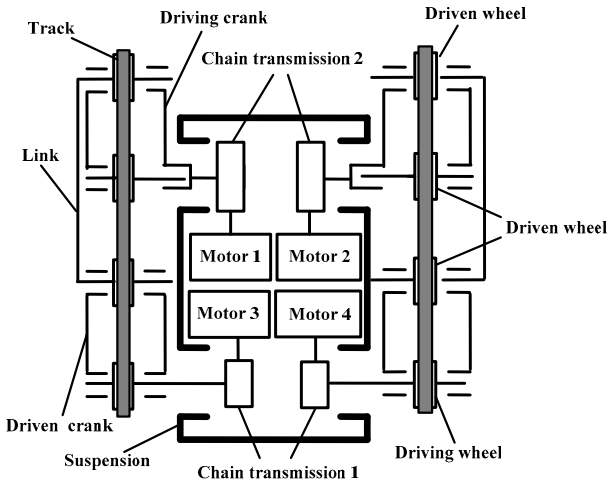
The prototype of RTMBot is constituted of a main body and a pair of symmetrical reconfigurable track modules based on a parallel four-bar mechanism as shown in

Fig.4(a). The outline dimensions of the prototype are 700mm long, 500mm wide and 300mm high. The structure of RTMBot is shown in Fig.4(b). The suspension, driving crank, driven crank and the link make up a parallel four-bar mechanism inside the reconfigurable track module.



(a) The prototype of RTMBot

(b) The mechanical structure of RTMBot



(c) The driving system of RTMBot

**Fig. 4.** The structure of RTMBot

The mechanism has 4 degrees of freedom as shown in Fig.4(c). Firstly motor 1 and motor 2 drive the two driving wheels on both sides through chain transmission 1. Then the left and right driving wheels can mesh with the tracks to keep transferring the movement. This is the reason that the robot can move forward, backward, and turn around at different rotational speeds of driving motors. The parallel four-bar mechanism configuration can be changed by motor 3 and motor 4 through chain transmission 2 and driving crank. It implements the reconfiguration of reconfigurable track module.

RTMBot has multiple locomotion configurations, which are realized by changing the configuration of reconfigurable track modules. Thus, the robot can work in complicated environments, such as steps, trenches, stairs and mono-obstacles, etc[15]. The robot can overcome the obstacles forward and backward, due to a symmetric structure along the body. This paper only analyzes the situation of trench-crossing forward, emphasizing on centroid-shifting caused by the reconfiguration.

### 2.2 Robot Centroid Configuration Model

The centroid configuration model of RTMBot is shown in Fig.2. In this figure,  $O_1x_1y_1 \sim O_5x_5y_5$  are respectively the coordinate systems of main body of RTMBot, driving crank, driven crank, link and track.  ${}^1T_2 \sim {}^1T_5$  are respectively the transformation matrixes from  $O_2x_2y_2 \sim O_5x_5y_5$  to  $O_1x_1y_1$ .  $\theta$  is the angle between driven crank and suspension at any position of the track during reconfiguration. Therefore, the coordinate matrix of RTMBot's centroid in  $O_1x_1y_1$  is defined below

$$\begin{aligned}
 {}^1P &= ({}^1P_1m_1 + {}^1T_2{}^2P_2m_2 + {}^1T_3{}^3P_3m_3 + {}^1T_4{}^4P_4m_4 + {}^1T_5{}^5P_5m_5) / m \\
 &= \begin{bmatrix} \frac{1}{m}(m_1{}^1x_1 + m_2{}^2x_1 + m_3{}^3x_1 + m_4{}^4x_1 + m_5{}^5x_1) \\ \frac{1}{m}(m_1{}^1y_1 + m_2{}^2y_1 + m_3{}^3y_1 + m_4{}^4y_1 + m_5{}^5y_1) \\ 0 \\ 1 \end{bmatrix} \tag{1}
 \end{aligned}$$

Where  $m_1 \sim m_5$  are respectively the mass of main body, driving crank, driven crank, link and track,  $m$  is the total mass of RTMBot,  ${}^iP_j$  is the coordinate matrix of  $m_i$  in  $O_jx_jy_j$ ,  ${}^ix_j$  and  ${}^iy_j$  are the coordinates of  $m_i$  in  $O_jx_jy_j$ .

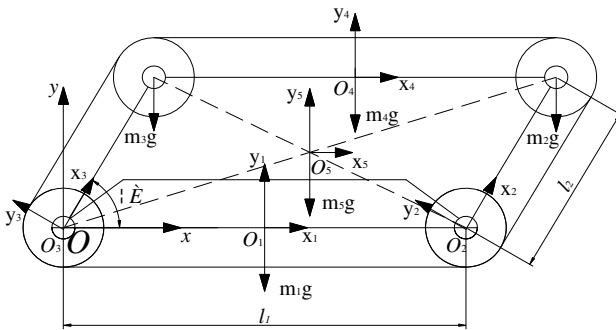


Fig. 5. The coordinate system of RTMBot

It is assumed that the centroids of driving crank and driven crank are located in the center of the driven wheels.  $l_1$  is the length of the suspension as well as the link.  $l_2$  is the length of driving crank as well as driven crank. The relations can be defined as

$$\begin{cases} {}^1x_1 = 0, {}^1y_1 = 0 \\ {}^2x_1 = l_1/2 + l_2 \cos \theta, {}^2y_1 = l_2 \sin \theta \\ {}^3x_1 = -l_1/2 + l_2 \cos \theta, {}^3y_1 = l_2 \sin \theta \\ {}^4x_1 = l_2 \cos \theta, {}^4y_1 = l_2 \sin \theta \\ {}^5x_1 = l_2/2 \cos \theta, {}^5y_1 = l_2/2 \sin \theta \end{cases} \quad (2)$$

Then the coordinate matrix of RTMBot's centroid in geodetic coordinate system  $O_0x_0y_0$  is

$${}^0P = {}^0T_1^1P = \text{rot}(z, \theta_0) \text{rot}(y, \psi_0) \text{rot}(x, \varphi_0) \text{trans}(P_x, P_y, P_z) {}^1P \quad (3)$$

Where  $\theta_0$ ,  $\psi_0$  and  $\varphi_0$  are the angles of pitch, roll and yaw of the robot, respectively.

Set the center of the last rear wheel contacting with the ground as the original point of the robot coordinate system  $Oxy$ . Thus, in  $Oxy$ ,  ${}^i y$  is equivalent to  ${}^i y_l$ , which is shown in Eq.(2). And  ${}^i x$  is defined in Eq.(4).

$$\begin{cases} {}^1x = l_1/2 \\ {}^2x = l_1 + l_2 \cos \theta \\ {}^3x = l_2 \cos \theta \\ {}^4x = l_1/2 + l_2 \cos \theta \\ {}^5x = l_2/2 \cos \theta + l_1/2 \end{cases} \quad (\theta \in [0, \pi)) \quad \begin{cases} {}^1x = l_1/2 - l_2 \cos \theta \\ {}^2x = l_1 \\ {}^3x = 0 \\ {}^4x = l_1/2 \\ {}^5x = l_1/2 - l_2/2 \cos \theta \end{cases} \quad (\theta \in [\pi, 2\pi)) \quad (4)$$

And the coordinate matrix of RTMBot's centroid in  $Oxy$  can be defined as

$$P = \begin{bmatrix} \frac{1}{m} (m_1 {}^1x + m_2 {}^2x + m_3 {}^3x + m_4 {}^4x + m_5 {}^5x) \\ \frac{1}{m} (m_1 {}^1y + m_2 {}^2y + m_3 {}^3y + m_4 {}^4y + m_5 {}^5y) \\ 0 \\ 1 \end{bmatrix} \quad (5)$$

It is assumed that there is no relative displacement between the robot and ground during the reconfiguration of RTMBot, and set the original points of  $O_0x_0y_0$  and  $O_1x_1y_1$  are concurrent. Substituting the design parameters into Eq.(1) and Eq.(5), the relation between  $\theta$  and the abscissa of RTMBot's centroid in  $O_0x_0y_0$  and  $Oxy$  is shown in Fig.3. The change of  $\theta$  has significant influence on the robot's centroid. The maximum abscissa change value of RTMBot's centroid in  $Oxy$  is  $\Delta x = 55.1\text{mm}$ . It means the robot can change its mass allocation by changes of  $\theta$ . So the locomotion configurations of trench-crossing can be planned to improve the robot's performance with this rule.

**Table 1.** The design parameters of RTMBot prototype

Symbol	$l_1$	$l_2$	$m_1$	$m_2$	$m_3$	$m_4$
Value	500mm	250mm	28kg	3.2kg	0.8kg	2kg

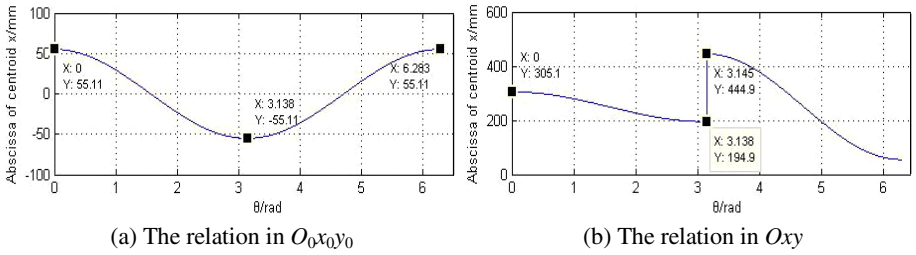


Fig. 6. The relation between  $\theta$  and the abscissa of RTMBot's centroid

### 3 Locomotion Plan and Calculation for Trench-Crossing

#### 3.1 Locomotion Configurations Plan

When the robot's centroid, the forepart and the rear-back contact with the ground are not in the trench at the same time, the robot can get across the trench successfully[3,4]. So the position of robot's centroid and the contact length between the foremost and the rearmost contact of the robot determine the maximum trench-crossing width. During trenches crossing, the contact length should be as long as possible; and the centroid of the robot should be as retral as possible during the beginning of crossing, and should be as anterior as possible during the end of crossing, so that the robot will get across trenches. RTMBot can reconfigure its shapes to change the position of its centroid. So we can use the RTMBot's centroid configuration model which is proposed above to analyse the reconfiguration of trench-crossing as shown in Fig.4. As a result, the RTMBot's trench-crossing performance can be improved. When  $\theta=0$  or  $\theta=\pi$ , the driving crack and the driven crack are in line, the contact length is maximum, which defined as the in-line configuration.

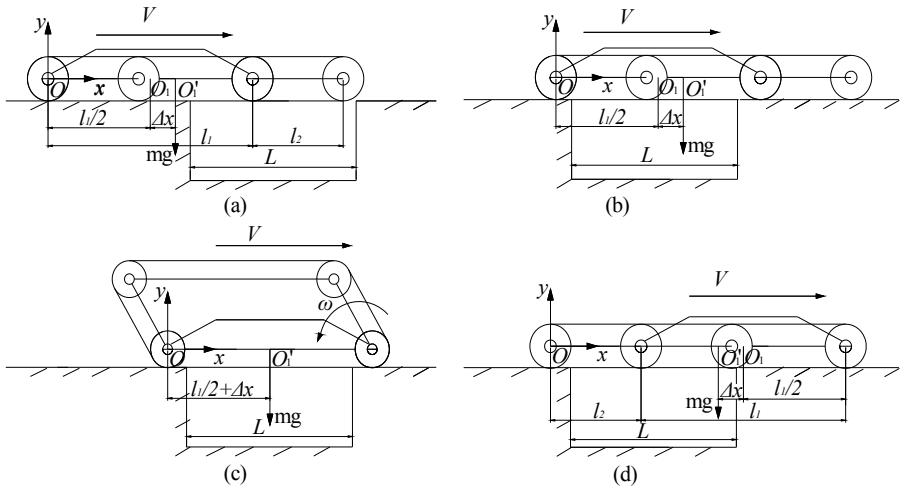


Fig. 7. The locomotion configurations during trenches crossing

When RTMBot confronted with a wide trench, as shown in Fig.4, set the center of the last rear wheel contacting with the ground as the original point of the robot coordinate system  $Oxy$  as mentioned above,  $O_1'$  is the centroid of whole RTMBot. At the beginning of the trench-crossing,  $\theta=0$ , the robot comes to the in-line configuration, and the centroid of RTMBot is retral (shown in Fig.4(a)). The abscissa of  $O_1'$  in  $Oxy$  is

$$x = l_1 / 2 + \Delta x \quad (6)$$

Then the robot gets across the trench. When the main body of the robot is in the middle of the trench as shown in Fig.4(b), the motor 3 and motor 4 rotate to make  $\theta=\pi$ , the robot comes to another in-line configuration as shown in Fig.4(c). Simultaneously, the centroid of RTMBot moves to the front of RTMBot. The abscissa of  $O_1'$  in  $Oxy$  is

$$x = l_2 + l_1 / 2 - \Delta x \quad (7)$$

After that, RTMBot moves forward to cross the trench as shown in Fig.4(d). In this process, the centroid, the foremost and the rearmost contact of RTMBot are not in the trench at the same time, so RTMBot could get across the trench successfully.

Then, assume RTMBot confronts with a narrow trench. It may not need the locomotion configurations as shown in Fig.4(b)~Fig.4(d). Make  $\theta=0$ , RTMBot comes to the in-line configuration as shown in Fig.4(a), and the robot could get across the narrow trench successfully.

### 3.2 Maximum Trench-Crossing Width Calculation

Through the analysis above, the locomotion configurations as shown in Fig.4 are applied to get the maximum trench-crossing width of RTMBot  $L_1$

$$L_1 = \min(l_1, \frac{l_1}{2} + l_2 - \Delta x) \quad (8)$$

If the locomotion configurations are not applied, the maximum trench-crossing width of RTMBot  $L_2$  is

$$L_2 = \min(\frac{l_1}{2} + \Delta x, \frac{l_1}{2} + l_2 - \Delta x) \quad (9)$$

Where  $l_1$  is the length of the suspension as well as the link,  $l_2$  is the length of driving crank as well as driven crank.  $\Delta x$  is the distance between the centroid of main body  $O_1$  and the RTMBot's centroid  $O_1'$  when  $\theta=0$  or  $\theta=\pi$ . The design parameters of RTMBot prototype are listed as follows:  $l_1=500\text{mm}$ ,  $l_2=250\text{mm}$ , and  $\Delta x=55.1\text{mm}$  according to Fig.3. Thus,  $L_1=444.9\text{mm}$ ,  $L_2=305.1\text{mm}$ , and the maximum trench-crossing width of RTMBot is  $L = L_1=444.9\text{mm}$  in theory.

## 4 Trench-Crossing Capability Experiments

### 4.1 Virtual Experiments

The software RecurDyn has been used for the analysis and simulation of RTMBot performance. The process of RTMBot crossing a trench with a width of 440mm is shown in Fig.5. Fig.6 shows the relative height of the robot's centroid.

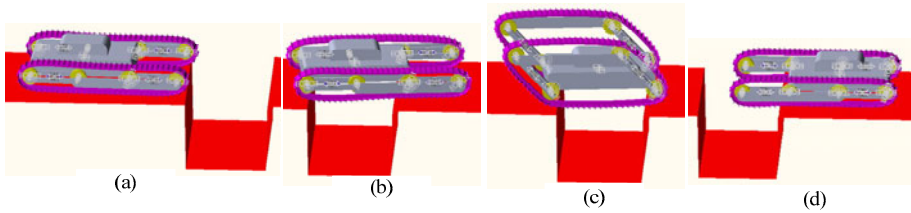


Fig. 8. Process of trench-crossing

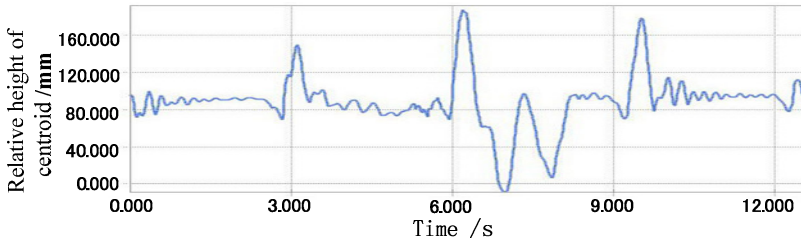


Fig. 9. The relative height of the robot's centroid

During the period of 0~3s, the robot came to in-line configuration. At the end of this period the front and second wheels of the robot reached the opposite side of the trench, and the rear wheels stayed on the original side as shown in Fig.5(b). After that, the robot reconfigured its shape, made  $\theta=\pi$  as shown in Fig.5(c).Then during the period of 7~10s, RTMBot moved forward, and the robot's centroid got across the trench. Finally the robot crossed the trench.

In another virtual experiment, without the locomotion configurations, the maximum trench-crossing width of the robot is 300mm.

### 4.2 Prototype and Experiments

According to the analysis and simulation of RTMBot, a reconfigurable tracked mobile robot prototype has been produced as shown in Fig.1(a). And two groups of experiments have been carried out to verify the design concept and its embodiment. Group 1 applied the locomotion configurations expressed above, while the other did not in contrast. The results are listed in Table 2.

Table 2. The maximum trench-crossing width of RTMBot

Groups	Theory analysis	Simulation	Prototype experiments
1	445mm	440mm	420mm
2	305mm	300mm	285mm

However, the data obtained from theory analysis, simulation and prototype experiments are consistent, which finally proves the correctness and feasibility of its structure and trench-crossing capability analysis.



The reasons for different results between theory analysis, simulation and prototype experiments can be illustrated as: (i) the centroid of the robot prototype laid behind the theoretical and virtual one; (ii) track slippage occurred during the reconfiguration; (iii) in order to prevent the robot from falling into trench, all the experiments were prudential.

Table 3 gives some other detailed parameters, which make clear that the RTMbot has a combined characteristic no worse than that of PackBot, especially for the trench-crossing capability.

**Table 3.** The performances of RTMBot

Robot quality	35kg
Dimensions	700mm×500mm×300mm
Maximum velocity	2.6m/s
Maximum load capacity	65kg
Maximum climbing angle	48°
Maximum step-crossing height	255mm
Maximum trench-crossing width	420mm
Maximum suspension-raising height	295mm
Maximum safe drop height	1.2m

## 5 Conclusions

A reconfigurable tracked mobile robot (RTMBot) has been proposed and its operational characteristics of trench-crossing have been investigated theoretically and practically. Some conclusions are drawn below:

(1) The robot structure is introduced. A centroid configuration model is proposed to analyse the trench-crossing capability.

(2) The trench-crossing locomotion configurations are planed and the maximum trench-crossing width of RTMBot is calculated.

(3) A virtual robot prototype has been constructed, and been used for performance analysis of the robot during trench-crossing.

(4) A robot prototype has been developed, and experiments have been carried out to prove the correctness and feasibility of the design concept and operation modes.

## Acknowledgment

This work was supported by the National University of Defense Technology (NUDT) graduate innovative fund under grant number S090301 (2009). The authors gratefully acknowledge the input of the anonymous reviewers.

## References

1. Junyao, G.: Survey of Military Mobile Robotics Systems in Foreign Country. *J. Robot.* 25(7), 746–755 (2003)
2. Dongqing, Y., Wangyong, W., Yue, M.: Status and Trends of Small Unmanned Ground Weapon Mobile Platforms. *J. Robot.* 26(4), 373–379 (2004)

3. Lee, K.M., Shah, D.K.: Kinematic Analysis of a Three-Degrees-of-Freedom In-parallel Actuated Manipulator. *J. IEEE Journal of Robotics and Automation* 4(3), 354–367 (1988)
4. Gosselin, C., Angeles, J.: The Optimum Kinematic Design of a Planar Three-Degree-of-Freedom Parallel Manipulator. *J. Journal of Mechanisms, Transmissions and Automation in Design* 110(3), 35–41 (1988)
5. Wells, P., Deguire, D.: TALON a Universal Unmanned Ground Vehicle Platform, Enabling the Mission to be the Focus. In: *Proc. SPIE*, vol. 5804, p. 747 (2005), doi:10.1117/12.602887
6. Halme, A., Leppanen, I., Salmi, Y., et al.: Development of Workpartner-Robot- Design of Actuating and Motion Control System. *J. Research Applies AI to Computer Animation* 7, 2–5 (2004)
7. Xingguang, D., Qiang, H., Jingtao, L.: Design and Implementation of a Small Ground Mobile Robot with Multi-locomotion Models. *J. China mechanical engineering* 18(1), 8–12 (2007)
8. Xingguang, D., Qiang, H., Kejie, L.: Design and Motion Analysis of Miniature Wheel-Track-Legged Mobile Robot. *J. Chinese Journal of Mechanical Engineering* 41(8), 108–114 (2005) (in Chinese)
9. Hutian, F., Yi, O., Xiaoyan, G.: Dynamics Modeling and Analysis of Miniature Ground-Moving Robot based on Special Pose. *J. Journal of Nanjing University of Science and Technology* 30(4), 486–490 (2006) (in Chinese)
10. Yamauch, I.B.: Packbot: a Versatile Platform for Military Robotics. In: *Proc. of SPIE on C. Unmanned Ground Vehicle Technology V I*, vol. 5422, pp. 228–237 (2004)
11. Jinneng, M., Xiaofan, L., Zhen, Y., et al.: Dynamic Modeling and Analysis for Obstacle Negotiation of Ground Mobile Robot. *J. Robot.* 30(3), 273–278 (2008)
12. Diansheng, C., Xi, Y., Tianmiao, W.: Kinematic Analysis and Control of Mobile Mini-Robots. *J. High Technology Letters* 18(1), 37–41 (2008) (in Chinese)
13. Kim, J., Lee, C., Kim, G.: Study of Machine Design for a Transformable Shape Single-tracked Vehicle System. *Mechanism and Machine Theory* 45, 1082–1095 (2010)
14. Govers, F.X.: Elbit Tracks&wheels VIPeR. *J. Robot Magazine*, 24–25 (2009)
15. Xiqing, Z.: Design and Research on the Deformabal Tracked Robot with Four Links. *National University of Defense Technology, D. Changsha* (2009) (in Chinese)

# Automatic Cooking Robot with a Novel Feeding System

Weixin Yan, Enguang Guan, Wentao Ma, Zhuang Fu, and Yanzheng Zhao

School of Mechanical Engineering, Shanghai Jiao Tong University, State Key Laboratory of Mechanical System and Vibration, Shanghai, P.R. China

{Weixin.Yan, Enguangovo, Wentao.Ma, Zhfu, Yzh-Zhao}@sjtu.edu.cn

**Abstract.** The cooking of traditional Chinese food is very complex. In order to finish the cooking efficiently, an automatic Chinese cooking robot with a novel feeding system is designed. The cooking materials can be poured into the boiler accurately and orderly, by the feeding system. The feeding system is composed of material box, transport mechanism and turnover mechanism. Controlled by the DSP system, the cooking robot can accomplish the feeding process automatically. Experiments show that the feeding system with a high performance system can satisfy the feeding requirements for Chinese Automatic cooking robot.

**Keywords:** automatic cooking robot; feeding system; DSP.

## 1 Introduction

Chinese traditional cooking has a long history and profundity in worldwide. But the cooking equipments available now can only make simple cooking processes such as heating with the microwave oven, baking with the roaster, boiling with an electric kettle, broiling with the frying pan. They can not complete the core part of the Chinese cooking processes such as pan-frying, stir-frying, burst-frying, quick-frying and re-frying with the related automated operations [1].

With the development of advanced science and technology, the research of automatic cooking robot meet the requirement of Chinese Cooking industrialization gradually [2]. And the high pressure of the modern life makes people spend less time on cooking, as well as the Chinese cuisine is very difficult. So it is urgent to design the robot to release people from cooking [3].

Feeding system is an important component of the automatic cooking robot. Working with the cooking robot, the automatic feeding system is a remarkable innovation and progress to the existing cooking equipments of the catering industry.

The current cooking robots can't accomplish the complicated technics of Chinese dishes. They just simulate the action of the human being and are commanded to finish some simple motion of cooking. Carnegie Mellon University designs a human-robot interaction system which could communicate with users. At first, a robot chef teaches

novices by providing detailed explanations of cooking tools. Then it adds incentives for speed and accuracy and replicated [4].

Yasushi Nakauchi designs a cooking support robot, which suggests what the human should do next by voice and gesture. Experimental results confirmed the feasibility of the inference system and the quality of support. They propose human activity recognition system, which infers the next human action by taking account of the past human behaviors observed so far [5-6].

In this paper, a Chinese cooking robot is introduced at first. It can learn the actions of high-class chef and simulate the whole process of cooking by control the several kinds of action modules and adjust the intensity and time of the fire. Then, the more important is, a novel feeding system is presented in this paper.

## 2 The Structure of Chinese Cooking Robot

Fig.1 shows the outlook of automatic cooking robot for Chinese dishes.



**Fig. 1.** The Chinese cooking robot

The automatic cooking robot for Chinese dishes has the following functions:

- 1 It can put the materials for cooking into the wok automatically.
- 2 It can make the food in the wok heated evenly.
- 3 It can finish the basic Chinese cooking technics.
- 4 It can clean the wok by itself.

The cooking robot includes several kinds of functional modules: the novel feeding system, the module of leaving the material out in the middle process, the cooking module, the fire-control module and man-machine communication module. The functions that can be finished by the modules are shown in Table 1 [7].

**Table 1.** The main functions for different modules

No.	Module	Function
1	Feeding system	Add the cooking materials into the wok accurately
2	The module of leaving the materials out in the middle of process	1) Pour Get the cooking materials out of the wok completely
		2) Drain the oil Extract water and oil from the cooking materials, treated after drained by oil.
		3) Back to the wok After the cooking materials been drained of oil, they will be put back to the wok
3	Cooking module	1) Even stirring Mix the ingredients and heat them evenly
		2) Stir-fry and disperse Separate the ingredients immediately and heat evenly.
		3) Gathers together Gather the ingredients into the bottom of the wok
		4) Turn another side invert the ingredients completely (Heating surface and relative non-heating surface inversion)
		5) Guards against sticks Prevent the conglutination of the ingredients with the wok in order not to be burnt
		6) Crack Crack the ingredients by heating into pieces
		7) Drench Let the stock cover or soak the ingredients
4	Fire-control module	Control the switch of cooking range and the size of fire
5	Man-machine communication module	1) Control every sub module
		2) Recode the process of cooking action
		3) Man-machine communication

The feeding system is one of the most important modules in Chinese cooking robot.

### 3 Design Requirements of the Feeding System

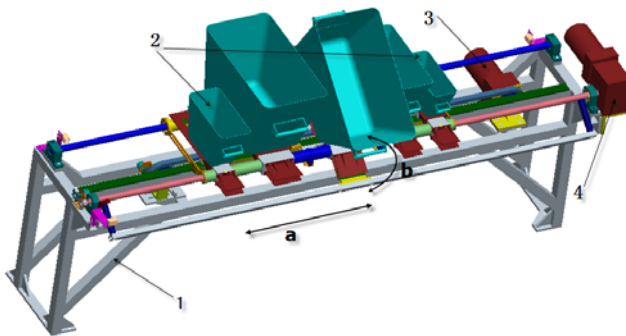
The action of feeding system is to put the cooking materials automatically. According to the cooking technics, materials could be classified as major ingredient, minor ingredient. The feeding system container should be designed to certain capacities by the dish classification while meeting the batch feeding action. Meanwhile, to satisfy the requirement of cooking technics, the feeding system should be designed flexibly. Thus, the feeding system could be used in different sorts of cooking robots.

There are two types of action requirement related to feeding system. One is manual action including loading materials, which is to pour the container full of raw materials onto the feed mechanism before cooking and get down the empty container. Another is automatic feeding action controlled by a control device, which demands to pour the raw materials into pot according to the cooking technics. This action should be taken into more attention while designing the feeding system.

## 4 Mechanical Design of Feeding System

Fig.2 shows the detailed information of the feeding system. The main function of the automatic feeding system is realized by the horizontal moving of the container components and the rotating of the single container. Motor No.3 manages the horizontal moving of the feeding system to let container component move straightly along feed rack. When one of the containers reach the feed place, motor No.4 manages the rotating feeding action to rotate single container in certain angle range. The feeding position which locates in the middle of the whole system is reasonable, as it has the smallest length in the horizontal path.

We utilize two kinds of sensor to accomplish the position detection of horizontal motion. The first sensor is hall magnet sensor which judge whether the container touch the initial position, the other sensor is absolute magnet encoder which can measure the distance of horizontal moving by the method of sending pulse. And we use two hall magnet sensors for the action of rotating, they are equipped at two ultimate position.



**Fig. 2.** The feeding system (1. Rack 2. Container components 3.Horizontal motor 4. Rotating feeding motor)

### 4.1 Container Components

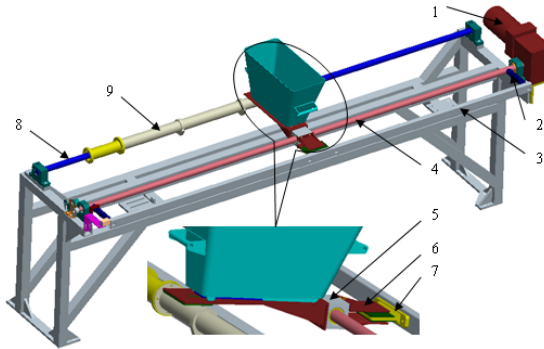
The feeding system use boxlike containers, there are ginger and garlic container, major ingredient container, two accessories containers and flavoring container from right to left. The arrangement meets general cooking feed order.

We use the push and pull mode to fetch and put containers. There exists a trapezoidal groove in the bottom of each container and a flashboard fixed in feeding system

which match the groove. This kind of structure could not only fix containers into rack, but also make them more convenient.

## 4.2 Rotating Components

We use this component to realize the rotating action of containers to put raw materials into pot when cooking. Taking major ingredient container for example, Fig.3 shows the feeding system and the rotating feed components and ignores those components which have little to do with the action.



**Fig. 3.** Rotating feeding system

The major action of the rotating feed is to drive the oriented axis rotating and the retroflex rod through the connecting rods at the edges of the axis. To let the containers move there fixed a board which insert into the gap of two tubes with some room on one side of the container on the retroflex rod. The board rotates by the retroflex rod, so the splints on the container move to make the whole container rotate with the retroflex base around the axis. The parts in the base are linear ball sleeves, which could move both straight and around. There is polyurethane cushion pad in the area of both sides of the board and the supporting board contacted with the tubes to prevent the metal collision noise when rotating.

## 4.3 Horizontal Moving Components

These components mainly control the containers moving to the feeding position in the middle of the device while cooking. To finish this work, the device needs to carry five independent containers moving forward and backward along the axis and the retroflex axis.

Fig.4 shows the related components in the system. Five containers are distributed between the spacing tubes and the assorted spacing tubes. Fig.4 shows the locations of a, b, c, d and e. The containers match the oriented retroflex axis by the linear ball tube in the base, as well as the supporting board is put on the tube of the oriented axis. Put the tube on the oriented axis with its edges built in linear bearing so that it could move horizontally. As the linear ball tube could move along the oriented retroflex axis well, all the containers could move horizontally.

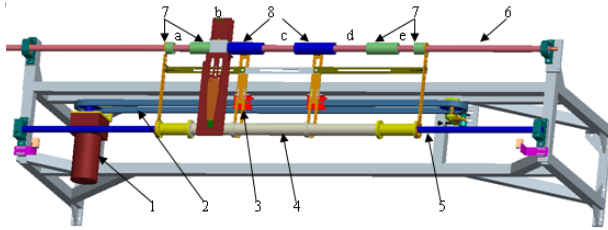


Fig. 4. Horizontal moving system

#### 4.4 Bending Deformation Analysis Based on ANSYS

The oriented axis and oriented retroflex axis shoulder nearly the whole weight of the feeding system mainly include five containers, raw materials, frame components and bearings. When the feeding system is heavy as the amount of food is large, the bending deformation of both axes effect directly the smoothness of linear gearing and linear ball tubes moving along axis. To relieve this bending, we could lighten the system, enlarge the diameter of axes, and shorten the distance of support saddle [4].

We design the feeding system in the cooking robot to output least 10kg dishes every time at each pan. The main body reaches 38kg and the distance between the axes is 1610mm, all the weight calculation depends on the Solid model computing function of PRO/ENGINEER [5] except raw materials and bearings weight. To ensure reasonable axis diameter, we choose finite element method, analyze structural static mechanics using ANSYS and take the degree of axes bending deformation as the reference condition of structure design.

According to the permitted deflection formula for general axis [3]:

$$y_{\max p} = (0.0003 \sim 0.0005) \times 1610 = (0.483 \sim 0.805) \text{mm}$$

The axis bending deformation is decided by load and its distribution. The analysis process is given as follows. First, we make sure the work position that may lead axis bending, then establish finite element model for each working state and by analyzing the deformation with ANSYS, we could get reasonable axes diameters and bending deformation. From the analyzing result, the largest deformation occurs when major ingredient container reaching feed position and not feeding yet; also, the deformation of oriented retroflex axis is a little larger than that of oriented axis. When pick diameters 25mm and 30mm, Fig.5 shows the deformation situation of oriented retroflex axis and the maximum value are 0.743mm and 0.385mm. So we could conclude that 25mm diameter could meet the basic requirement, and 30mm is more stabilized.

In order to increase the stiffness of the feeding system, we could relieve the axes load as well as picking large diameter axes. Fig.6 shows the supporting part structure of the system. We fix a set of supporting components whose mainly part is a rolling bearing rotating around the axis on the supporting rack on five containers' boards. When not rotating feeding, the rolling bearings contact with a supporting girder of the rack and avoid making a noise with the help of polyurethane cushion pads. When axes are undertaking load, most weight of containers and raw materials are transferred into the girder of the mounting rack. So the structure shares much load of axes to void large bending deformation.



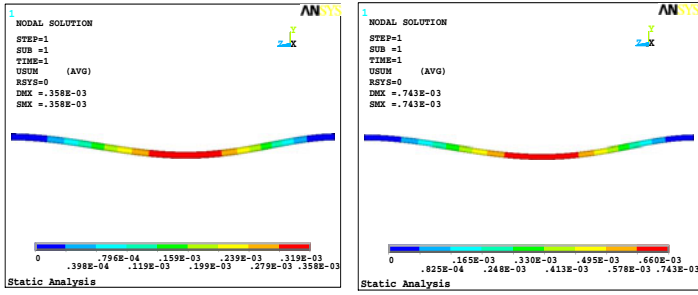


Fig. 5. The bending deformation of 25mm and 30mm diameters oriented retroflex axes

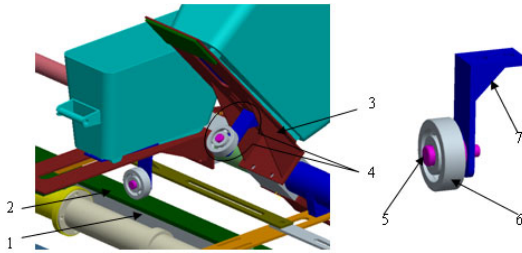


Fig. 6. The supporting parts of the feeding system

## 5 Control System for Feeding Systems

We choose TMS320LF2407A DSP of Texas Instruments to be the control hardware. We need mainly control the translatory and rolling DC motors requiring the speed control; the encoders and the proximity switches sending the feedback of position signal to the DSP; the two-way current analog inputs are got to the DSP, too. The upper layer touch screen offers friendly human-computer interaction interface, controls feed system and communicate with the DSP by CAN bus. Fig.7 shows the control system structure.

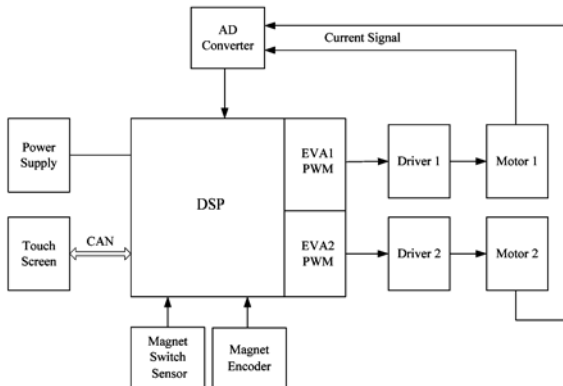


Fig. 7. Structure of feed control system

### 5.1 Motor Control

There are two event manager modules in the DSP. Every comparator, universal timers, dead-time unit and output logic could produce a pair of programmable dead time and polar PWM output on two special Electronic device pins [6]. Each event manager module owns 6 channel PWM outputs which could offer 6 PWM signals directly controlling the motor drives. 12 PWM signals from two event manager modules could in charge of 6 DC motors, which is far too enough to manage the transferring and rolling motors. And we could achieve the purpose of controlling motors and adjusting the speed by set up relevant registers.

### 5.2 The Communication of DSP and Host Computer

TMS320LF2407A has CAN (Controller Area Network) control module, and it could communicate with host computer by CAN bus. The CAN information frame consists of frame start, Arbitration field, control field, data field, check field and frame end as well as supports standard and extended format. We define the communication protocol as below [7]:

A. The information frame includes 8 bits data domain which could be divided into data frame and command frame. The former stands for the specific action of device such as motor action; the later means the demand information from host computer to DSP such as system reset.

B. the DSP and the host computer will shake hands every time before consequent communication and then host computer sent detailed action order and parameters. While feeding, the DSP would supply the current system state to host computer so that the later could display and diagnose it.

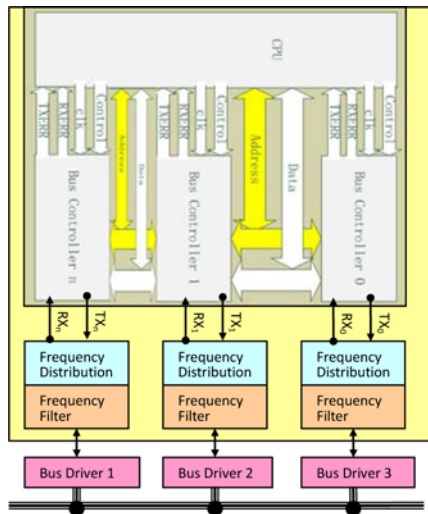


Fig. 8. The structure of redundancy FCS IP core

In this paper, we use FPGA (Field Programmable Gate Array) to make the redundancy FCS (Fieldbus Control System) of the automatic cooking robot come true, Nios II IP core is embedded in the system. Compared with the traditional industrial control technique, this system which uses redundancy FCS has the character of high reliability and applicability. The Fig.8 shows the structure of redundancy FCS IP core in FPGA. The Using of redundancy FCS can promote the development of the control technology of the robots [8].

### 5.3 DSP Software Flow

The DSP is the center of feeding system control. Fig.9 shows the software flow chart. In the chart the system state self-check mainly examine whether components locates in proper place and the DC motor locked or blocked. Once this occurs, the device stop running and host computer displays the possible failure cause.

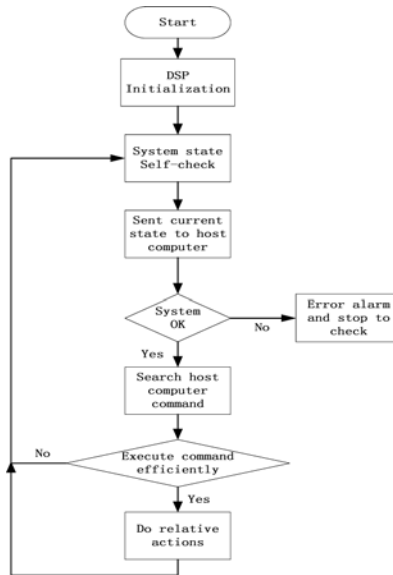


Fig. 9. DSP program flow chart



Fig. 10. Feeding system of stir-frying machine

## 6 Experiments

According to the design above, we equip the final feed system after assembled into the automatic cooking robot (stir-frying machine) which could cook large amount food each time which is shown in Fig.10. The automatic cooking robot with the feeding system in this paper has developed many Chinese dishes.

In the experiments the automatic cooking robot (stir-frying machine) has cooked the traditional Chinese dish whose name is fried shredded pork with green peppers. This dish's major raw ingredient is shredded pork and its minor ingredient is green pepper. It requires the strict time of feeding and accurate intensity of fire.



**Fig. 11.** The moment of pouring the material

Fig.11 shows the action of pouring the material. In order to pour the material into pot completely, we designed some special actions to assist feeding. The first one is to make the container rap the frame several times when make the rotating action, then the remnant material would drop into the pot by the reverse force. The second one is to control the acceleration of container rotation. To control the current of motor can

**Table 2.** Accuracy ratio of feeding

Index	Major material	Minor material	Ginger	Seasoning
1	96%	98%	95%	86%
2	97%	98%	94%	90%
3	97%	97%	96%	88%
4	98%	98%	94%	88%
5	97%	97%	98%	92%
6	95%	99%	92%	87%
7	96%	96%	92%	91%
8	95%	97%	93%	90%
9	96%	96%	92%	86%
10	97%	97%	94%	86%
Average	96.4%	97.3%	94.0%	88.4%

control the torque of motor, and then control the acceleration of motor, so we design the current feedback circuit to finish this work.

We have finished feeding test ten times as table 2. The feeding system can almost pour the major raw material, minor material and ginger completely. Because of the viscosity of the seasoning, the hit ratio of seasoning can reach above 86%, the hit ratio of the major material can reach above 95%.

## 7 Conclusion

As the most important module of the cooking robot, the feeding system realizes the functions of feeding raw materials according to Chinese cuisine technics while cooking. Successfully horizontal moving and retroflex feeding, chronically running of no breakdown, are meet the design requirements perfectly. Experiments show that the feeding system has good flexibility to fit many sorts of material feed requirements. And the automatic cooking robot with the feeding system has a potential utility for the better life in future.

**Acknowledgments.** This work was partially supported by the Hi-Tech Research and Development program of China (863 project, under Grant No. 2007AA041601, 2007AA041602 and 2007AA041603).

## References

1. Yan, W.X., Fu, Z., Liu, Y.H., Zhao, Y.Z., Zhou, X.Y., Tang, J.H., Liu, X.Y.: A novel automatic cooking robot for Chinese dishes. *J. Robotica*, 1–6 (2007)
2. Bin, H.: The trend of industrialized development for Chinese cuisine and its principles. *J. Culinary Science Journal of Yangzhou University*, 20–70 (2003)
3. Zhang, Y.-s.: Analyze the industrialization of Chinese cuisine. *J. Chinese cuisine research* 1.2, 50–54 (1996)
4. Torrey, C., Powers, A., Marge, M., Fussell, S.R., Kiesler, S.: Effects of Adaptive Robot Dialogue on Information Exchange and Social Relations. In: *HRI 2006* (2006)
5. Nakauchi, Y., Fukuda, T., Noguchi, K., Matsubara, T.: Time Sequence Data Mining for Cooking Support Robot. In: *2005 IEEE International Symposium on Computational Intelligence in Robotics and Automation* (2005)
6. Fukuda, T., Nakauchi, Y., Noguchi, K., Matsubara, T.: Human Behavior Recognition for Cooking Support Robot. In: *2004 IEEE International Workshop on Robot and Human Interactive Communication Kurashiki, Okayama* (2004)
7. Robert Bosch GmbH. *CAN Specification Version 2.0* (1991)
8. Yan, W.X., Yu, Q.X., Fu, Z., Li, Y., Zhao, Y.Z., Zhou, X.Y., Tang, J.H., Liu, X.Y.: High Reliability Embedded Fieldbus Control System for Cooking Robot. In: *2nd International Symposium on Robot and Artificial Intelligence, Tokyo* (2008)

# Multi-sensor Based Autonomous Underwater Manipulator Grasp

Zhihu Xiao, Guohua Xu, Fuyuan Peng,  
Guo yuan Tan, Xiaolong Xu, Xiong Shen, and BoYang

Underwater Technology Laboratory, College of Navy and Ship Science and  
Engineering, Huazhong University of Science and Technology, 430074  
Wuhan, Hubei, China  
zhihuxiao@163.com

**Abstract.** As the development of mine, oil and gas resources exploitation in the ocean, the underwater manipulators become necessary tools. But operation of manipulator is tedious. To reduce its dependence on operator, autonomous manipulation is studied. With the support of the Chinese high technology develop program (the 863 program), we design an underwater 3500m electrical driven manipulator. It is equipped with a unique ultra-sonic probe array and an underwater camera. Autonomous grasping a cylinder using ultra-sonic sensor and vision is studied. The sensor system and strategy of autonomous manipulation are discussed in details. Experiments results demonstrated the manipulator autonomous work ability.

**Keywords:** underwater, manipulator, multi-sensor, autonomous grasp.

## 1 Introduction

Ocean is paid more and more attention today, so the underwater vehicle with manipulator is regarded as one of the most important tools in ocean exploring. In most of commercial work-class remote operate vehicles (ROVs) today, the manipulator is operated by an operator. The work performance depends on operator's skill. Remote operation is tedious and hard work. It is lack of vision telepresence. Lots of study has been done in order to reduce the burden of operator. Supervisory control [1] is often used. A vision servo system [2] was studied in Twin-Burger 2. The manipulator can sample jelly fish with the help of stereo vision servo. A new working space control system[3] was developed in the Korea Ocean Research & Development Institute (KORDI) ROV Hemire in order to assist operator during coring, drilling and underwater connector mating. Semi autonomous underwater vehicle for intervention mission (SAUVIM) [4] has been developed in Hawaii University. Today, it is one of the first underwater vehicles (if not the only one) capable of autonomous manipulation. It can retrieve hazardous objects. It is one of the few examples worldwide of applications using autonomous manipulation for underwater intervention technology. Although these researches are done, underwater autonomous manipulation in a hazardous and unstructured environment isn't appropriately solved.

The aim of this paper is to study manipulator autonomous manipulation. In this paper, the target is a cylinder with artificial square in surface. The manipulator is equipped with a unique ultra-sonic probe array and an underwater camera. We design a four steps strategy (searching, alignment, reaching and grasping) in whole process. Ultra-sonic sensor is to obtain distance and approximate location. Vision is used for recognize feature of the cylinder. Both of vision and ultra-sonic sensor are applied in alignment target. The structure of the paper is given as following. First part is overview of the whole system, then sensor system, the grasp strategy. Some experiments of autonomous grasp are given at last.

## 2 System Overview

The underwater manipulator (named huahai-4E, stands for four functions deep ocean electric manipulator in China) is an electric manipulator designed to catch some objects in deep ocean. The underwater manipulator is designed to be mounted in the bottom of the underwater vehicle (ROV or AUV). It features open frame architecture, integrated joints which is oil filled to compensate high water pressure. It is of three degree of freedoms (DoFs). They are shoulder revolving, shoulder rising/ down, elbow rising /down respectively. It is design to be mounted in the bottom of the underwater vehicle (ROV or AUV). Fig. 1 shows the system application concept and the manipulator DH coordinate system. Ultra-sonic sensors, vision camera are fixed in the manipulator arm. In autonomous mode, with the help of sensors, it can catch target objects autonomously.

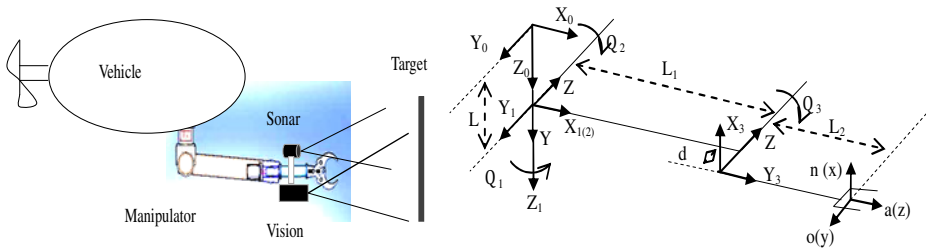


Fig. 1. System overview and manipulator D-H coordinate system

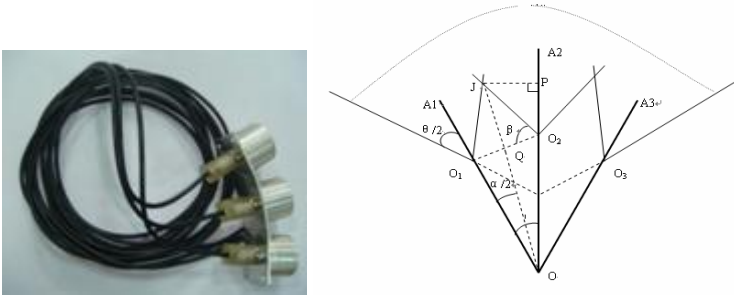
## 3 Ultra-Sonic Sensor and Vision

### 3.1 Ultra-Sonic Array

Underwater acoustic are widely used in underwater mapping and target detection. its detection mechanism is emitting sound pulses and detecting or measuring their return after sound wave being reflected.

Ordinary, one ultra-sonic probe is enough to measure the distance of target, so one probe is often used. Kleeman[5] , LeMay[6] introduced multi-sonar system. In our design, a unique three ultra-sonic probe array is designed. These three probes are

arranged in a convex surface. It is shown in fig. 2. The advantage of the arrangement is that it can obtain a larger detectable angle, and detect both distance and approximate location of object.



**Fig. 2.** Ultra-sonic probe array and wave beam angle

In Fig.2, the A1, A2, A3 are the axes of three probes respectively.  $\alpha$  is the angle between the two axes. Wave beam angle of single probe is  $\theta$ . As wave beam is axial symmetry. The angle from the axes of probe to its wave beam boundary is  $\theta/2$ . The left and right boundaries are the detection limits. The total equivalent wave beam angle is  $(\theta+2\alpha)$ . Wave beam angle of single probe is  $11^\circ$ , so probe array can covers  $30^\circ$  through convex surface arrangement.

The ultra-sonic sensor system is of three main units: micro computer unit (MCU, msp430), sending circuit and receiving circuit. The core unit (msp430) manages time sequence of sending and receiving, switch control of the three ultra-sonic probes. In a work period, every probe emits  $8\mu s$  500K Hz pulses and receives reflect wave once. MSP430 calculates the distance of target, and figures out its approximate location.

### 3.2 Underwater Vision

In order to grasp an object autonomously, determining the target object is the first problem. SIFT (Scale Invariant Feature Transform) was developed by Lowe [7] for image feature generation in object recognition applications. The features are invariant to image translation, scaling, rotation, and partially invariant to illumination changes and affine or 3D projection. These characteristics make it widely applied in mobile robot. In our design, The SIFT features are chosen for target recognition. The target is a cylinder with artificial square in surface which is shown in Fig.3.

Lots of researchers studied on monocular vision for camera positioning, McLauchlan[8] introduce an variable state-dimension filter (VSDF) recursive method Azarbayejani[9] introduce Recursive estimation of motion, structure, and focal length. Davison presented Mono SLAM algorithm [10].

In this paper, the manipulator can only grasp target in vertical direction. The distance information can be obtained by ultra-sonic sensor, so different with ordinary application, the target orientation and position isn't calculated by vision data.



Direction of feature point in horizon plane of camera reference frame is calculated. The optic axis of camera is parallel with axis of arm. The centroid of cylinder is chosen as the feature point.  $P(X_{cm} \ Y_{cm})$  is the centroid of cylinder in image plane.

In ideal pin-hole camera model, the effective focal length ( $f$ ) is considered. So the projection transformation is:

$$X_u = (x/z) \cdot f \quad Y_u = (y/z) \cdot f \quad (1)$$

The coordinates of the point is  $P(x, y, z)$ . In image plane, coordinates of the point is  $P_u(X_u, Y_u)$ .  $P_w(x_w, y_w, z_w)$  is one point in world coordinate system. In camera reference frame, the coordinates of the point is  $P(x, y, z)$ . In image plane, coordinates of the point is  $P_u(X_u, Y_u)$ .  $P_u(X_u, Y_u)$  is the image coordinates in the idea pin-hole camera model.

From the equation 1,

$$\tan(\alpha) = X_{cm} / f \quad (2)$$

$\alpha$  is direction angle of centroid of cylinder in horizon plane of camera reference frame. The centroid of cylinder is calculated by cylinder boundary line which is detected though Hough Transform after edge detection. It is shown in Fig.3.

Because of flow, impurity in water, strengthening glass in front of lens, the image is distorted. Image nonlinear correction is necessary. Considering the distortion, the following nonlinear model has been adopted. In distortion model, the new normalized point coordinate  $P_d$  is defined as follows:

$$P_d = \begin{bmatrix} X_d(1) \\ X_d(2) \end{bmatrix} = (1 + kc(1) \times r^2 + kc(2) \times r^4 + kc(5) \times r^6) P_u + Dx \quad (3)$$

Where  $Dx$  is the tangential distortion vector :

$$Dx = \begin{bmatrix} 2kc(3)X_uY_u + kc(4)(r^2 + 2X_u^2) \\ kc(3)(r^2 + 2Y_u^2) + 2kc(4)X_uY_u \end{bmatrix} \quad r^2 = X_u^2 + Y_u^2$$

The undetermined parameter  $Kc$  is a distortion vector,  $Kc = [kc(1) \dots kc(5)]^T$ ;  $P_m(X_m, Y_m)$  is the image coordinates on the image plane.

$$P_m = [X_m \ Y_m \ 1]^T = K[X_d(1) \ X_d(2) \ 1]^T \quad (4)$$

where  $K = \begin{bmatrix} fc(1) & \alpha_c \times fc(1) & cc(1) \\ 0 & fc(2) & cc(2) \\ 0 & 0 & 1 \end{bmatrix}$ .

$[cc(1), cc(2)]$  is the point of intersection of optical axis and image planar.  $\alpha_c$  is skew coefficient.  $[fc(1), fc(2)]$  is the focal distance expressed in units of horizontal and vertical pixels. Bouguet method [11] has been adopted in underwater camera calibration, therefore  $[cc(1), cc(2)]$ ,  $\alpha_c$ ,  $[fc(1), fc(2)]$  and  $Kc$  can be determined.

$P_d$  is the new coordinates after nonlinear correction.

$$\tan(\beta) = X_d(1) / fc(1) . \tag{5}$$

$\beta$  is the direction angle of the cylinder's centroid in horizon plane of camera reference frame after nonlinear correction.

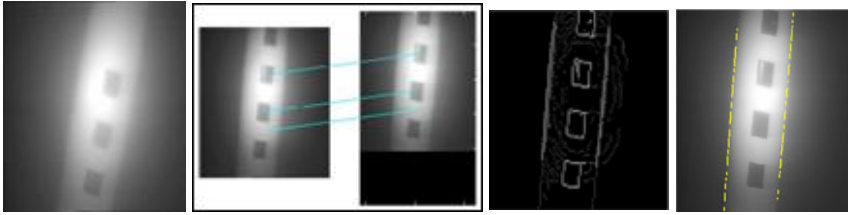


Fig. 3. Target, target recognition using SIFT, edge detection, line detection

### 4 Autonomous Grasp Strategy

There are four steps in whole grasp schedule. They are searching, alignment, reaching and grasping. In joint space, the motion is planed. Q1, Q2, and Q3 are angle of shoulder revolving joint, shoulder rising/ down joint, and elbow rising /down joint respectively.

Without a global view camera, so the manipulator is hard to know where the target is. Searching the target is necessary at first step. The manipulator elbow is in horizon orientation when the manipulator searches in its whole workspace. The searching trajectory plan is:

- (a)  $Q1=0^\circ$  to  $180^\circ$      $Q2=90^\circ$      $Q3= -90^\circ$  ,
- (b)  $Q1= 180^\circ$      $Q2=90^\circ - \theta_k$      $Q3=-90^\circ + \theta_k$  ,
- (c)  $Q1= 180^\circ$  to  $0^\circ$      $Q2=90^\circ - \theta_k$      $Q3=-90^\circ + \theta_k$  ,
- (d)  $Q1=0^\circ$      $Q2=90^\circ - \theta_k$      $Q3=-90^\circ + \theta_k$  ,
- (e)  $Q1=0^\circ$  to  $180^\circ$      $Q2=90^\circ - \theta_k$      $Q3=-90^\circ + \theta_k$  ,
- (f) Do (b) until the target is found in vision or time out.

$\theta_k$  is a interval angle. Its value should be chosen appropriately. If it is too small, the total search time will be long, but if it is too big, the manipulator maybe couldn't find the target in high speed motion.

In alignment step, our purpose is to make sure the target cylinder is right ahead of manipulator jaw. Both of ultra-sonic and vision is applied in this step. A target location factor  $f_u$  of ultra-sonic sensor is used. It is shown in Tab.1.

**Table 1.** Ultra-sonic sensor target location factor

probe Target Location	Left Probe	Middle Probe	Right Probe	factor $f_u$
Left	$L_L$	0	0	-1
Left of centre	$L_L$	$L_m$	0	-0.5
Right Ahead	0	$L_u$	0	0S
Right of centre	0	$L_m$	$L_R$	0.5
Right	0	0	$L_R$	1
No Target Found	0	0	0	0

The alignment trajectory plan is

$$(a) Q1(t+1) = Q1(t) + \beta K_1 + f_u K_2 \quad Q2(t+1) = Q2(t) \quad Q3(t+1) = Q3(t).$$

(b) Do (a) until  $\beta < \beta_m$  and  $f_u = 0$  or time out.

$K_1 K_2$  are two weight factors which can be chosen in experiment.  $\beta$  is calculated through equation 5.  $\beta_m$  is a small threshold.  $\beta_m$  should be small enough and make sure the target right ahead of manipulator jaw.

In reaching step, the manipulator reaches toward the cylinder target.  $L_u$  is the distance of target.  $L_u$  and data of the joints angle sensor are used for calculating the plan joints angle.

(a) Keep Q1 unchanged; calculate Q1 Q2 through  $L_u$  and joint sensor data;

(b) Do (a) until  $L_u - L_{fix} \in [0 \quad R]$

$L_{fix}$  is the distance from the center of jaw to ultra-sonic middle probe,  $R$  is the diameter of jaw when jaw is open.

In the last step: Driven by the jaw motor, the jaw accomplishes open or close function. Torque sensor in jaw joint is applied to end grasp.

## 5 Autonomous Grasp Test

In order to test the autonomous ability, experiment is established. The diameter of target cylinder is 45mm. The following pictures show grasp tests in ship tank.



Fig. 4. Underwater autonomous grasp tests

The joints angle in the whole procedure is given in Fig.5.

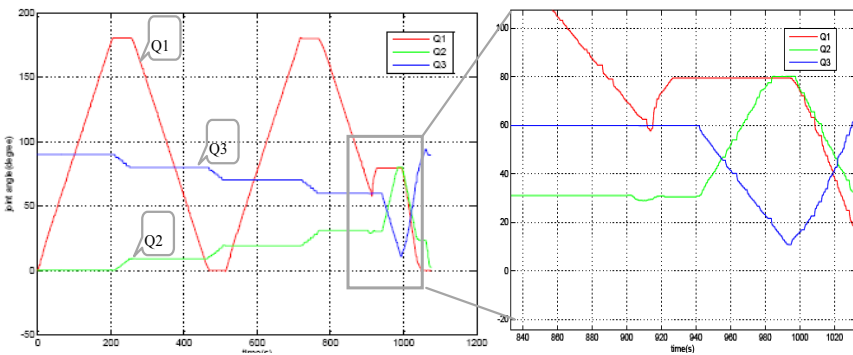


Fig. 5. Joints angle (Q1, Q2, and Q3) value while grasping

In Fig.5, from 0 to 900 seconds, the manipulator searched the target in its workspace. Then it found the target. To align the jaw with target, it took almost 20 seconds. Later, it reached toward target. After 990s, it returned to its original place.

In experiments, we found the range of ultra-sonic sensor is two meters. Because of attenuation of sound wave, when the target distance increases to more than one meter, the reliability and precision decreases. Vision recognition ratio is about 80%. The processing frequency is more than 1.25Hz. The installation error and the measure error of the ultra-sonic array and the camera result in grasping failure.

## Conclusions

An underwater manipulator autonomous grasping a cylinder target is presented. The manipulator is equipped with a unique ultra-sonic probe array and an underwater camera. The ultra-sonic sensor is used for obtaining distance and approximate location of target. Vision has been adopted in recognition of the cylinder that has a pattern of black squares in surface. Both of vision and ultra-sonic sensor are used for alignment jaw with the target. There are searching, alignment, reaching and grasping four steps in whole process. Experiments results demonstrated that manipulator can grasp target without operator’s assistance. There are some improvements left to be done in

future. For example the target cylinder is static and with artificial feature. Autonomous grasp unstructured objects in more complex environment will be studied.

## Acknowledgment

This work was funded by the China National 863 Program “deep-sea manipulator key technique” project (2006AA09Z203), State Commission of Science and Technology for National Defense Industry “micro underwater work tool” project and National Science Foundation of China (Grant No. 50909046).

## References

1. Yoerger, D., Slotine, J.J.: Supervisory control architecture for underwater teleoperation. In: IEEE Int. Conf. on Robotics and Automation, ICRA 1987, pp. 2068–2073 (1987)
2. Yann, E., Nose, Y., Ura, T.: Autonomous underwater sampling using a manipulator and stereo visualservoing. In: Oceans 2005 - Europe, OCEANS 2005, Brest, vol. 2, pp. 731–736 (2005)
3. Jun, B.H., Shim, H.W., Lee, P.M., Baek, H., Cho, S.K., Kim, D.J.: Workspace control system of underwater tele-operated manipulators on ROVs. In: OCEANS 2009-EUROPE, OCEANS 2009, Bremen, pp. 1–6 (2009)
4. Marani, G., Choi, S.K., Yuh, J.: Underwater autonomous manipulation for intervention missions AUVs. *Ocean Engineering* 1, 15–23 (2009)
5. Kleeman, L., Kuc, R.: Mobile Robot Sonar for Target Localization and Classification. *Int. J. of Robotics Research* 4, 295–318 (1995)
6. Lemay, J.B.L.J.: Error Minimization and Redundancy Management for a Three Dimensional Ultrasonic Ranging System. In: Proc. of the IEEE/RJS Int. Conf. on Intelligent Robots and Systems, Raleigh NC, pp. 837–844 (1992)
7. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Key points. *Int. J. of Computer Vision* 2, 91–110 (2004)
8. Mclauchlan, P.F., Murray, D.W.: A Unifying Framework for Structure and Motion Recovery from Image Sequences. In: Proc. ICCV 1995, pp. 314–320 (1995)
9. Azarbayejani, A., Pentland, A.P.: Recursive Estimation of Motion, Structure and Focal Length. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 562–575 (1995)
10. Davison, A.J.: Real-Time Simultaneous Localization and Mapping with a Single Camera. In: IEEE Int. Conf. on Computer Vision, pp. 1403–1410 (2003)
11. Bouguet, J.Y.: Visual Methods for three-dimensional Modeling. California Institute of Technology Pasadena, California (1999)

# The Targets Pursuit for Multi-robot System with Hybrid Wireless Sensor Networks

Kun Shi, Zhiqiang Cao, Wenwen Zhang, and Chao Zhou

Laboratory of Complex Systems and Intelligence Science, Institute of Automation,  
Chinese Academy of Sciences, Beijing 100190, China  
{kun.shi, zhiqiang.cao, wenwen.zhang, chao.zhou}@ia.ac.cn

**Abstract.** This paper gives a targets pursuit approach for multi-robot system in a multi-region workspace with the help of hybrid wireless sensor networks, which comprises static sensor nodes and mobile sensor nodes. The layout of hybrid wireless sensor networks are firstly presented, and the static nodes network adopts multi-target tracking based on dynamic cluster with the constraints of K-level and communication cost. Considering the hollow zones and multiple-region division, cooperative target tracking with both static and mobile sensor nodes as well as multi-region joint target tracking is designed. Based on the obtained targets information, the pursuit task is executed by the robots, which is verified by the simulations.

**Keywords:** Multi-robot system, Targets pursuit, Hybrid wireless sensor networks, Multi-region workspace, Hollow zone.

## 1 Introduction

Multi-robot System (MRS) has received many attentions from robotic researchers with the characteristics of distributed, parallelism, robustness, etc [1] [2]. Considering the limitations of onboard sensors of robotic system, the wireless sensor networks (WSN) provides a good solution. Recently, the combination of robotic system and WSN, as a kind of network robot systems (NRS) [3], has becomes a hotspot. Some NRS systems have been presented. The URUS project [4] designs a network of robots that in a cooperative way interacts with human beings and the environment. PEIS ecology project [5] aims at providing cognitive and physical assistance to people. NRS project in Japan [6] is to develop systems that provide robot services by combining robot and network technologies. Liu and Yang introduce a ubiquitous and cooperative service framework for NRS [7].

In this paper, we conduct the multi-robot pursuit with the help of hybrid wireless sensor networks. A regional-level network is considered and each region has a centered regional gateway, which is responsible for its targets management and communications with the robots or other gateways within the communication range. There is a user layer responsible for the collection and processing of information from the regional gateways. For a region, limited by the number of static sensor nodes, there maybe exist hollow zones, which require the mobile sensor nodes. Therefore, the

target is tracked by the hybrid wireless sensor networks with static sensor nodes and mobile sensor nodes. According to the targets in the environment, the corresponding robots are dispatched to execute the pursuit task.

The rest of the paper is organized as follows. In Section 2, the multi-robot pursuit with hybrid wireless sensor networks is described in detail. The layout of hybrid wireless sensor networks, multi-target tracking and robots pursuit are presented. Simulation results are demonstrated in Section 3 and Section 4 concludes the paper.

## 2 Multi-robot Pursuit with Hybrid Wireless Sensor Networks

The whole system comprises static sensor nodes, mobile sensor nodes, regional gateways, user layer and robots. The environment is divided into some non-overlapped regions with equal area. For simplicity, the static/mobile sensor nodes are called static node and mobile node for short, respectively.

Each region includes a number of static nodes whose layout adopts K-level coverage constraint and each node maintains a regional perception coverage map. When the static nodes within the region can't coverage full area, the hollow zones will be considered. In this case, suitable mobile nodes are sent for dynamic patrol, which requires a reasonable patrol path. For static nodes tracking, a dynamic cluster-based target tracking with the constraints of K-level and communication cost is adopted. When a target is confirmed by a node, if there is not a dynamic tracking cluster for this target, an initial cluster with a leader and several members will be formed. During the tracking process, the leader will synthesize the target information from all member nodes to acquire the position of the target by hierarchical multi-target data association. Meanwhile, the tracking cluster is updated for better tracking.

For the mobile node, it patrols according to the pre-set path. Once it finds a target which is not tracked by other nodes, it will apply and become a leader candidate of the target. When the target moves from the static zone into the hollow zone or conversely, the target is cooperatively tracked by the static nodes and mobile nodes. When the target moves into another region, multi-region joint tracking is used. The event trigger scheme is adopted to trigger the communication between the regional gateways and the user layer. When the event of *a new target appearing*, *target cross-region* or *target being caught* happens, the targets information from the regional gateways is sent to the user layer, which will unify the targets information and deliver them to regional gateways. Based on the information of targets and robots, it executes the task allocation and scheduling, and then multiple robots execute the task.

### 2.1 The Layout of Hybrid Wireless Sensor Networks

The layout of sensor nodes is the basis of the wireless sensor networks. For each node, it has certain sensing range. Limited by the nodes' number and the environment to be monitored, a reasonable layout is important, which may cover a larger area with fewer nodes. In this paper, the layout in a regular way [8] with K-level coverage constraint is used to find the proper sensor node interval  $d_s$ , which makes the most situations can be perceived by K nodes. Then static nodes will be deployed counter-clockwise from the outside to the inside. After the layout of static nodes completes, according to hollow zones, suitable mobile nodes are sent to the zones to patrol.

Take the region  $z_j$  with the size of  $d_c \times d_r$ , for example. In order to avoid the blind spots of perception in the hollow zone, the mobile node is required to monitor the zone.

Perception coverage of region  $z_j$  is described by the grid. The region may be divided into a grid map with the size of  $l_r \times l_c$  and each grid has a binary variable, which describes whether the grid can be detected. If the center of the grid is located in the sensor node's perception range, the grid is considered to be perceived and recorded as a covering grid with the value of 1, otherwise, it will be recorded as a hollow grid. After the regional perception coverage map has been generated, the hollow zone will be extracted by the linked hollow grids.

For each effective hollow zone, the first step is to select the patrol path points according to the principle of greater coverage with fewer points. The whole hollow zone is required to be divided into some sub-regions with equal area. For each sub-region, select a point to ensure that the sub-region may be covered when the mobile node is in the point. For simplicity, the sub-region is designed to the closest shape to quadrilateral inscribed to the perception circle with the radius  $R_s$  of the sensor node. If the area of a sub-region is proper, its centroid will be selected as a path point.

After the path points for hollow zone are obtained, it is required to plan the patrol path to visit all the points with less time. Assume that the mobile node is set to move in a constant speed. The problem may be abstracted as a weighted graph. A patrol path is a Hamilton loop with returning back to the starting point, which is a typical TSP problem [9][10]. In this paper, we use 2-opt based local search method to obtain the path. That is, for a legal path, replace the two sides with the other two non-path edges. If the cost after the replacement is less than that of original path, the new path is chosen until a proper path is acquired. When the mobile node does not find a target, it will patrol according to the patrol path.

As shown in Fig. 1, the environment is divided into 4 regions. It is seen that the region 2 and region 3 each has a hollow zone with 4 and 8 path points, respectively. The corresponding patrol paths are also given.

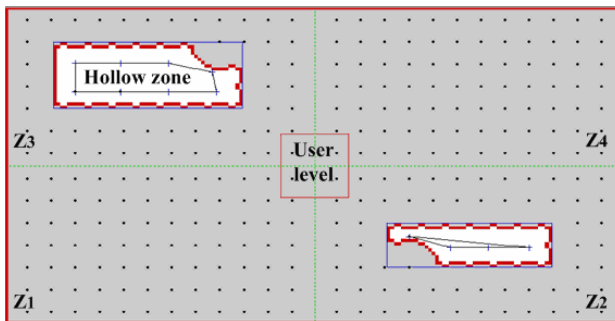


Fig. 1. The hollow zones and patrol paths of mobile nodes

### 2.2 Multi-target Tracking

In this section, a multi-target tracking approach based on dynamic cluster with communication cost constraint and the K-level constraints is described. Communication



cost of sensor nodes includes the cases of data sending and receiving, and the communication cost [11][12] may be expressed as follows:

$$C = P_T + P_R = (p_T + p_p d^\alpha) m_T + p_R m_R, \tag{1}$$

where  $p_T$  and  $p_R$  are the costs of sending/receiving unit information, respectively; and  $p_p$  is the path consumption of transferring unit information per unit distance;  $\alpha$  is path consumption factor;  $m_T$  and  $m_R$  are the amount of sent/received information, respectively.

We denote with  $S_t^t = (P_t, U_t)$  the state vector of target in time  $t$ , where  $P_t = (\tilde{x}_t, \tilde{y}_t, \tilde{\theta}_t)$  is the pose vector of target,  $U_t = (\tilde{v}_t, \tilde{\omega}_t)$  reflects the motion parameters of target, and  $\tilde{v}_t$ ,  $\tilde{\omega}_t$  are the linear velocity and angular velocity of target, respectively. Therefore, the prediction to the target is shown as follows.

$$P_t = P_{t-1} + \begin{bmatrix} \tilde{v}_t \times \cos(\theta_{t-1}) \times \sin(\tilde{\omega}_t T) - \frac{\tilde{v}_t}{\tilde{\omega}_t} \times \sin(\theta_{t-1}) \times [1 - \cos(\tilde{\omega}_t T)] \\ \tilde{\omega}_t \\ \tilde{v}_t \times \cos(\theta_{t-1}) \times [1 - \cos(\tilde{\omega}_t T)] + \frac{\tilde{v}_t}{\tilde{\omega}_t} \times \sin(\theta_{t-1}) \times \sin(\tilde{\omega}_t T) \\ \tilde{\omega}_t T \end{bmatrix} \tag{2}$$

where  $T$  is the sampling interval.

### 2.2.1 Target Tracking Based on Dynamic Cluster for Static Nodes

Dynamic cluster based target tracking with K-level constraint is to select K nodes with less communication cost to compose a cluster for target tracking. There are a leader node and several member nodes in a tracking cluster. The leader node is responsible for information collection from member nodes, target state estimation, and cluster update according to the K-level constraint and communication cost constraint.

In a tracking cluster with K-level constraint, the average communication cost for a leader is  $C_k = \sum_{i=1}^k C_{iL} + \frac{C_L + C_0}{T_{\min}^k}$ , where  $k$  is the nodes number of cluster (including the leader and members);  $C_{iL}$  is the communication cost of the  $i$ th member and the cluster leader and  $C_{iL} = (p_R + p_T + p_p d_{iL}^\alpha) \cdot m_c$ , where  $m_c$  is amount of measurement information;  $C_L$  is the communication cost of the new leader and the old leader and  $C_L = (p_R + p_T + p_p d_{LL}^\alpha) m_G$ ;  $C_0$  is the fixed communication cost when leader changes, including cost of leader information broadcasting and transmission of target information, and  $C_0 = (p_T + p_R R_c^\alpha) m_0$ , where  $R_c$  is the communication range of the leader;  $T_{\min}^k = \min(T_i | i = 1 \dots k)$  is the max tracking time of  $k$  nodes.

The dynamic cluster based tracking approach with K-level constraint and communication cost constraint includes the formation of initial cluster and cluster update.

1) *The formation of initial tracking cluster*

When a new target appears, there is no leader node initially and each sensor node monitors the environment independently. Considering the environmental disturbance and the uncertainty of measurement values, for the node that detects the target, it needs to confirm whether this is a target according to the accumulated measurements. After a new target is confirmed, the first step is to form a unique initial tracking cluster for one target.

For a sensor node, it will send leader request information including the leader request symbol, its own serial number and confirmed target positions set to the surrounding sensor nodes. Meanwhile, it receives the leader request information from other nodes. If the sensor node becomes the leader by processing, it sets its targets state table. For each target, the leader node selects the member nodes according to the target position and K-level constraint, and notifies them. After the leader node broadcast its leader information, the nodes within its communication range will receive the information.

If a sensor node confirms a target, it first inquiries the related leader for the corresponding target state based on its stored leader table. If the target has been tracked by a leader node, it tags the target.

2) *Hierarchical multi-target data association*

Leader node is responsible for the state estimation in target tracking and the selection of new leader and members in proper time according to target state. As the leader may track multiple targets at the same time, the data association is required. For a leader node, we denote with  $\Omega_T^t = \{\omega_1, \omega_2 \cdots \omega_{N_t^t}\}$  the targets set at time t, where  $N_t^t$  is the number of tracking targets. The measurements at time t are divided into  $N_t^t + 1$  categories and  $\omega_0$  represents non-target class. Therefore, the targets set is extended to  $\Omega_T^t = \{\omega_0, \omega_1, \cdots, \omega_{N_t^t}\}$ . Priori state vector of target  $\omega_i (\omega_i \in \Omega_T^t)$  at time t is labeled as  $\hat{S}_i^t = [\hat{P}_i^t, \hat{U}_i^t]$ , where  $\hat{P}_i^t = (\hat{x}_i^t, \hat{y}_i^t, \hat{\theta}_i^t)$  and  $\hat{U}_i^t = (\hat{v}_i^t, \hat{a}_i^t)$ . All priori states vector is depicted as  $\hat{S}_t^t = [\hat{S}_1^t, \cdots, \hat{S}_{N_t^t}^t]$ . The data association process is shown as follows [13][14][15]: after the data association based on the maximum posterior distribution for sensor node is executed, considering that the target's prior position and sensor node's measurement have uncertainty, the optimal association within the cluster is required. After we may obtain the measurements corresponding to each target, by synthesizing them, the target information is confirmed.

3) *Update of tracking cluster*

With the continuous motion of the target, the cluster will have to be updated by the leader, which will select the better leader and member nodes to form a new cluster. In this paper, new leader and its members are selected based on K-level and communication cost constraints.

Firstly, the sensor nodes set  $C_T = \{(S_i, t_i, d_i) \mid \|S_i P_T\| < R_s\}$  that can detect the target is determined according to the current state of target  $T$ , where  $|C_T|$  is the elements number in  $C_T$ ,  $t_i$  is the transit time for  $S_i$  (the transit time is the estimated duration of the target within the perception area of a node), and  $d_i$  is the distance between  $S_i$  and the target. When the number of nodes that can detect the target does not satisfy the  $K$ -level constraint, all nodes will be involved in tracking, otherwise, communication cost matrix  $EM = \{em_{ij} \mid (S_i, S_j \in C_T)\}$  will be used. Each item  $em_{ij}$  denotes the average communication cost for leader node  $S_i$  and member node  $S_j$  and  $em_{ij} = (p_R + p_T + p_p d_{ij}^\alpha) \cdot m_c + \frac{1}{K} \cdot \frac{C_{iL} + C_0}{\min(t_i, t_j)}$ , where  $C_{iL}$  is the communication cost between leader node  $S_i$  and latest tracking cluster leader. Especially,  $em_{ii} = 0$ . Therefore, for the leader  $S_i$ , select  $K$  nodes with smaller  $em_{iv} (S_v \in C_T)$  to form the cluster for less communication cost. We denote with  $E_i$  the average cost. For the whole tracking system, the tracking cluster with minimum  $E_i$  is chosen as the new one.

### 2.2.2 Joint Target Tracking with Static Nodes and Mobile Nodes

Since there maybe hollow zones in the whole sensor network monitoring area, the mobile node is indispensable and it will track the target in the hollow zone to ensure that target is not lost.

Case 1: new target appearing in hollow zone

During the monitoring process of mobile node, when a target is confirmed, it judges whether to request for becoming a leader. If the mobile node itself becomes the leader or the leader is a static node, it will track the target. When the leader is another mobile node, it patrols.

Case 2: target moving from static zone to hollow zone

The leader node estimates the target motion according to the target information and the stored perception coverage map. When the target is about to enter a hollow zone, it calculates the entry point, and sends the moving tracking request to the regional gateway. When the gateway receives the information, it selects the mobile node with patrol state which is nearest to the entry point for this target. Then the mobile node will move toward the entry point.

After the mobile node becomes the candidate leader node, when it detects a target, it will match measured values with the target according to acquired leader information and the target state. Once the match is successful, the confirmed target is the target to be taken over. When it is nearer to the target, it sends take-over signal to the leader and thus becomes the new leader. The previous cluster will give up tracking.

If the state of the target changes without entering the hollow zone, the leader will send the cancel request to the regional gateway, which thus will release the mobile node. Thus the mobile node restores to the patrol state.

### Case 3: target moving from hollow zone to static zone

When the target moves in the hollow zone, a mobile node is responsible for tracking and estimating the target's state. When the target moves into the static zone, as the mobile node is limited in the hollow zone, thus it needs to release its duty of leader according to target state and the distribution of static nodes nearby the target in proper time. Static leader and member nodes will be selected in accordance with K-level constraint and communication cost constraint by mobile node, which then starts patrolling.

### 2.2.3 Multi-region Joint Target Tracking

To avoid the target be regarded as a new target when it moves from one region into another region, multi-region joint target tracking is required. In the target tracking process, when the target is estimated to leave the region according to the target motion and the stored regional boundary information, the leader sends a message to its gateway. After the gateway receives this message, it calculates the region the target will enter into, and sent the related information to that regional gateway, which will store the cross-region information. When a new target appears in that region, the target is matched with the stored information. If the match is failed, the target is considered as a new target, or else, the *target cross-region* event is triggered.

## 2.3 Robots Pursuit

After the targets to be pursued are provided by the hybrid sensor network, the multi-robot system may execute the pursuit of the targets. Firstly, user-level synthesizes the targets information from different gateways to assign a serial number for each target and establishes the association with the target number of the regional gateways, which will be sent to the regional gateways. The task allocation and dispatch are carried out. The process is depicted as follows:

When the event of a *new target appearing*, *target cross-region* or *target being caught* happens, the tasks are allocated. We denote with  $\Omega_T = \{\omega_1, \omega_2 \cdots \omega_{N'_T}\}$  the targets set, where  $N'_T$  is the number of targets and robots set is described by  $\Omega_R = \{R_1, R_2 \cdots R_{NR}\}$ , where  $NR$  is the robots number. The principle of allocation is to make the distances between the robots and the targets be shortest on the assumption that each target is pursued by no more than one robot. Define the distance matrix  $D$  with the size of  $NR * N'_T$  and each item  $d_{ij}$  denotes the distance of the robot  $R_i$  and the target  $\omega_j$ . Give an evaluation function  $d = \sum_{i \in [1, NR]} d_{i a_i}, d_{i a_i} \in D$ , where  $d_{i a_i}$  is the distance of robot  $R_i$  and its only corresponding target  $\omega_{a_i}$ . If the robot  $R_i$  has no target,  $d_{i a_i} = 0$ . The optimal allocation is to choose the robot-target pair with minimum  $d$ .

After the robot determine its target according to the allocation results of user level, if it can not obtain the information from the regional gateway in which the target locates, it means that the robot is far away from the region and it should directly move towards the regional gateway. Once the robot establishes the connection with the

gateway, based on the target inquiry result from the gateway, it will directly move towards the target for pursuit. When the robot is within a certain distance from the target, the target is considered to be captured.

### 3 Simulations

The whole environment is divided into 4 regions with equal area. The simulation parameters are as follows:  $d_c = 450$ ,  $d_r = 250$ ,  $l_c = l_r = 5$ ,  $R_s = 40$ ,  $d_s = 35$ ,  $K = 4$ . In target tracking,  $R_c = 100$ ,  $p_R = p_T = 20$ ,  $p_p = 2$ ,  $\alpha = 2$ . The linear velocities of robot and target are  $v = 12$  and  $v = 10$ , respectively.

Simulation 1 is used to test the situation of joint target tracking of static nodes and mobile nodes when the target moves through the hollow zone. Take region 3 as an example (shown in Fig. 2). During the target's movement process, when it is about to enter the hollow zone, the regional gateway selects suitable mobile node. When the mobile node is nearer to this target, it becomes the leader to track the target. After the target moves out of hollow zone, the appropriate static leader and member nodes are chosen for continuous tracking.

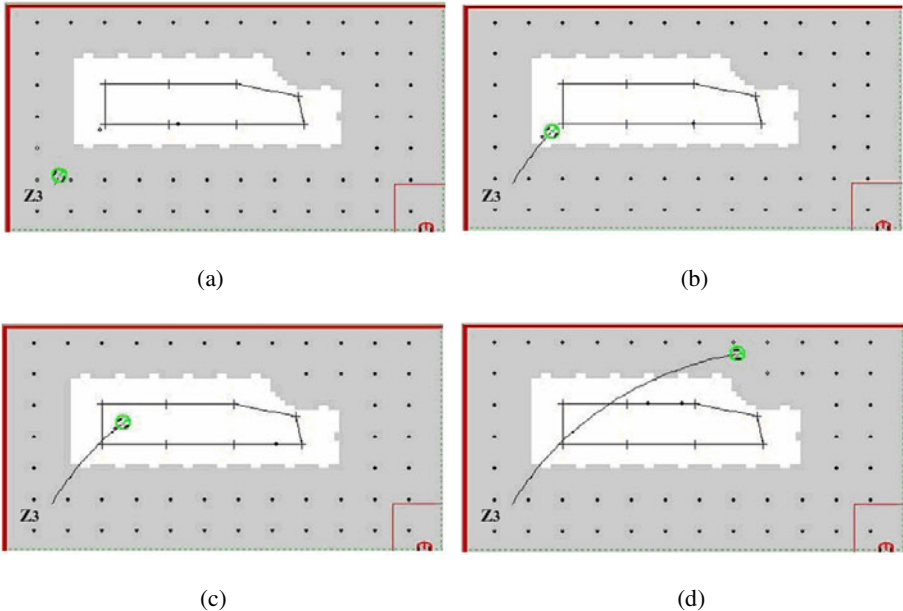


Fig. 2. Result of simulation 1

Simulation 2 testifies the multi-region joint tracking situation and the result is shown in Fig. 3. It is seen that the target is successfully tracked.

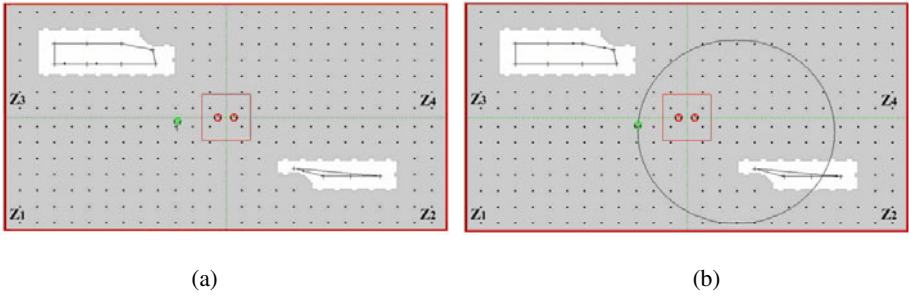


Fig. 3. Result of simulation 2

Simulation 3 gives the process of the multi-robot/multi-target pursuit, which is shown in Fig. 4. Initially, there are two targets  $W_1, W_2$  and two robots  $R_1, R_2$  are appointed to pursue them respectively. After the robot  $R_2$  finishes its task, a new target  $W_3$  is added. The task is re-allocated. The target of robot  $R_2$  is  $W_3$  and the target of the robot  $R_1$  remains unchanged. Finally, all targets are successful captured.

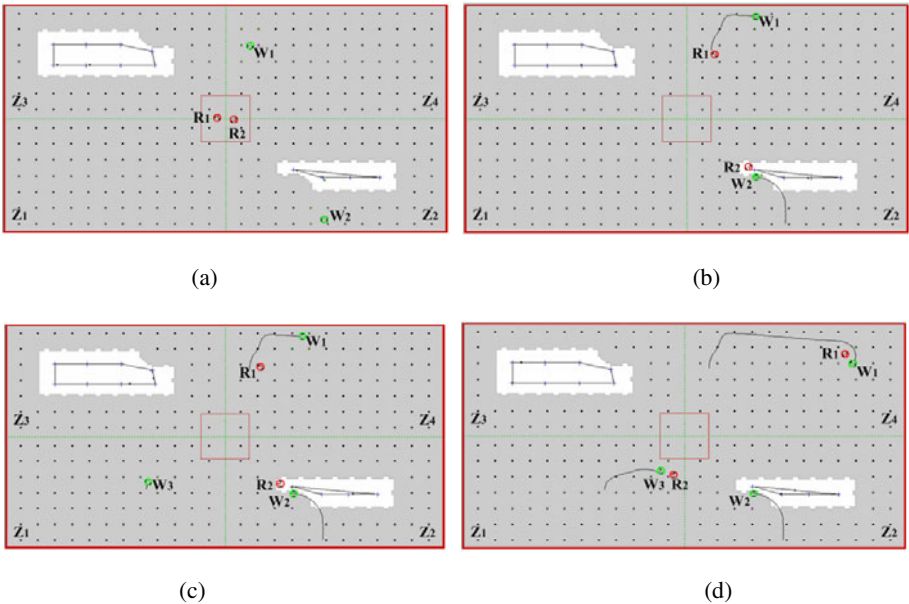


Fig. 4. Result of simulation 3

## 4 Conclusions

This paper conducts the research on targets pursuit of multi-robot system with the targets information provided by the hybrid wireless sensor networks. From the simulations

we have conducted, the targets are tracked effectively and the pursuit task is completed smoothly by the combination of static nodes and mobile nodes of multiple regions. The ongoing and future work includes the designs of the collision-free pursuit control for mobile robots as well as the active escaping strategy for the targets.

**Acknowledgments.** This work is funded in part by the National Natural Science Foundation of China under Grants 60805038, 60725309, and in part by the National High Technology Research and Development Program of China (863 Program) under Grants 2009AA043901-1, 2006AA04Z258.

## References

1. Cao, Y.U., Fukunaga, A.S., Kahng, A.B., Meng, F.: Cooperative mobile robotics: antecedents and directions. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 226–234 (1995)
2. Tan, M., Wang, S., Cao, Z.Q.: Multi-robot Systems. Tsinghua University Press, Beijing (2005) (in Chinese)
3. Sanfeliu, A., Hagita, N., Saffiotti, A.: Network robot systems. *Robotics and Autonomous Systems* 56(10), 793–797 (2008)
4. <http://urus.upc.es/index.html>
5. [http://aass.oru.se/%7epeis/frameset\\_page.html](http://aass.oru.se/%7epeis/frameset_page.html)
6. [http://www.irc.atr.jp/en/research\\_project/net\\_robo/](http://www.irc.atr.jp/en/research_project/net_robo/)
7. Liu, Y.B., Yang, J.H.: A Ubiquitous and Cooperative Service Framework for Network Robot System. In: Xie, M., Xiong, Y., Xiong, C., Liu, H., Hu, Z. (eds.) ICIRA 2009. LNCS (LNAI), vol. 5928, pp. 1–10. Springer, Heidelberg (2009)
8. Huang, C.F., Tseng, Y.C.: The Coverage Problem in Wireless Sensor Network. *Mobile Network Applications* 10(4), 519–528 (2005)
9. Wan, Y.Y., Zhou, Z., Chen, G.L., Gu, J.: SizeScale: New Algorithms for the Traveling Salesman Problem. *Journal of Computer Research and Development* 39(10), 1294–1302 (2002) (in Chinese)
10. Sun, H.L., Liu, Q.S., Hu, S.W.: Research on dynamic order insert crossover operator for travel salesman problem. *Journal of Computer Applications* 27(3), 633–634 (2007) (in Chinese)
11. Lindseya, S., Raghavendra, C.S.: Energy efficient all-to-all broadcasting for situation awareness in wireless ad hoc networks. *Journal of Parallel and Distributed Computing* 63(1), 15–21 (2003)
12. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-Efficient Communication Protocol for Wireless Microsensor Networks. In: Proceedings of the 33rd Hawaii International Conference on System Sciences (2000)
13. Oh, S., Russell, S., Sastry, S.: Markov Chain Monte Carlo Data Association for General Multiple-Target Tracking Problems. In: 43rd IEEE Conference on Decision and Control, pp. 735–742 (2004)
14. Bergman, N., Doucet, A.: Markov chain monte carlo data association for target tracking. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 705–708 (2000)
15. Zhang, W.W.: Research on Multi-robot Hunting and Distributed Target Tracking. M.E. thesis, Institute of Automation Chinese Academy of Sciences (2010) (in Chinese)

# A Fast Robot Path Planning Algorithm Based on Image Thinning

Ke Xu<sup>1,\*</sup>, Ruiqing Fu<sup>1</sup>, Lei Deng<sup>1</sup>, Yongsheng Ou<sup>1,2</sup>, and Xinyu Wu<sup>1,2</sup>

<sup>1</sup> Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences  
Shenzhen, China

<sup>2</sup> The Chinese University of Hong Kong  
ke.xu@sub.siat.ac.cn

**Abstract.** A fast, reliable mapping and path planning method is essential to autonomous mobile robots. Popular global algorithms usually adopt occupancy grid maps (OGM) or a topological map (TM) to generate a practicable path. In this paper, we present a new path planning algorithm based on the image thinning process and the extraction of key nodes. It requires much less time and memory compared with the OGM method for generating a path, and much simpler, more robust than the TM, while preserving advantages of both. With the new algorithm, a collision-free path could be obtained in real time. Results from simulation experiments validate its efficiency and reliability.

**Keywords:** mobile robots; thinning algorithm; path planning.

## 1 Introduction

Mapping and path planning are important for the navigation of mobile robots, especially in a complex environment. As a complex terrain is concerned, proper mapping and path planning methods could ensure robots move to the target quickly along a collision-free path. Popular path planning methods with known environmental information include occupancy grid maps (OGM) searching, topological map (TM) method and visible graph (VG) searching [1, 3, 4]. However, an OGM method is limited in real-time use since it splits the configuration space into large numbers of grids, which take a long time to compute. And an OGM could not restrain strong environmental noise [2]. VG searching represents obstacles with polygons. Paths from VG searching are always near edges of a polygon, which is likely to cause collision. The TM abstracts the environmental information to reduce computation and memory, but the procedure of TM algorithm is too complicated to implement. Besides, a TM is difficult to apply to accurate navigation because it lacks geometric information [5, 6].

In this paper, firstly we propose an improved thinning algorithm, by using which we get the single-connected thinning map (SCTM). A SCTM guarantees an effective and robust path planning for mobile robots. After that, we present a key node generation algorithm, which simplifies the SCTM to a key node graph (KNG) that preserves all information for navigation using much fewer graph nodes. On the basis of KNG, we apply A\* algorithm to generate a path. Results from simulation experiments



show that our method is superior to some traditional algorithms in both efficiency and planning effects.

## 2 Construction of the Single-Connected Thinning Map (SCTM)

This paper focuses on the construction of the SCTM from an OGM that could be constructed in real time. In an OGM, we use '1' to denote an occupied cell and '0' an empty cell.

Figure 1 illustrates a typical OGM where white grids denote obstacles. A SCTM is defined as the skeleton of an OGM, and it preserves the connectivity of structure of the OGM with the fewest grids. Such skeleton is an appropriate navigation path for mobile robots since it preserves the structure of the environment and is far away from surrounding obstacles.

Figure 2 shows the SCTM of Figure 1. Note that it completely maintains the connectivity information of Figure 1.

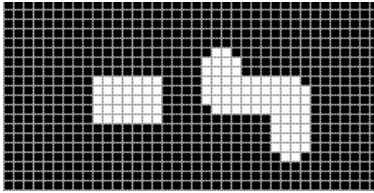


Fig. 1. An OGM

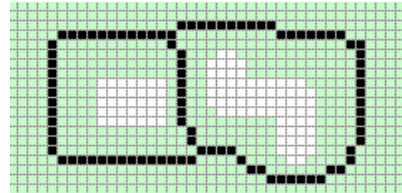


Fig. 2. The SCTM of Fig. 1

The construction of a SCTM is described as follows:

P <sub>9</sub>	P <sub>2</sub>	P <sub>3</sub>
P <sub>8</sub>	P <sub>1</sub>	P <sub>4</sub>
P <sub>7</sub>	P <sub>6</sub>	P <sub>5</sub>

Fig. 3. Center cell P<sub>1</sub> and its neighbors

Figure 3 illustrates the center cell under consideration (P<sub>1</sub>) and its eight neighboring cells (P<sub>2</sub>~P<sub>9</sub>). In the following discussion, '1' denotes an occupied cell and '0' an empty cell. In the thinning algorithm, we preserve each occupied cell and try to reduce the number of empty cells. Before applying the algorithm, we should firstly extend each obstacle area in the map by the radius of the robot. After such dilation, we could consider the robot as a cell in the map. We present the whole process of the generation of SCTM in Algorithm 1.

**Algorithm 1.** Single-Connected Thinning Map Generation

**Input:** An occupancy grid map  $M_1$  in which ‘0’ denotes an empty cell and ‘1’ an occupied cell

**Output:** The SCTM of  $M_1$

Dilating the obstacles using proper distance

**while** any cell changes from 1 to 0

**do**

**for** each cell in the map

**do if**  $2 \leq N(P_1) \leq 6$  and  $S(P_1) = 1$

**and**  $P_2 \cdot P_4 \cdot P_6 = 0$  and  $P_4 \cdot P_6 \cdot P_8 = 0$

**then**  $P_1 \leftarrow 0$

**if**  $2 \leq N(P_1) \leq 6$  and  $S(P_1) = 1$

**and**  $P_2 \cdot P_4 \cdot P_8 = 0$  and  $P_2 \cdot P_6 \cdot P_8 = 0$

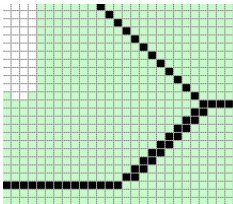
**then**  $P_1 \leftarrow 0$

**for** each cell in the map

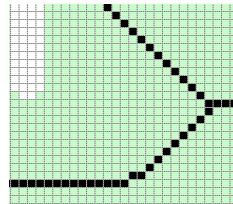
**do if**  $(P_3 \cdot P_4 \cdot P_6 = 1$  and  $P_9 = 0)$  or  $(P_4 \cdot P_6 \cdot P_7 = 1$  and  $P_9 = 0)$

**then**  $P_1 \leftarrow 0$

In Algorithm 1,  $N(p_1)$  denotes the number of cells that are not zero and  $S(p_1)$  denotes the number of changes from 0 to 1 in the sequence of  $p_2, p_3, \dots, p_8, p_9, p_2$ . The generation of a SCTM is based on the image thinning algorithm [7]. Traditional image-thinning algorithms could not ensure that the final skeleton is single-connected, which means for each cell  $C_1$ , there is at most one cell that links  $C_1$  to another edge. Figure 4 shows a part of a thinning map. Note that abundant cells are still preserved. According to robot path planning, such cells would bring unnecessary computation and duplicate paths. Here we improve the traditional algorithm so we could eliminate them as Figure 5 shows. Each edge in the map is transferred into a single-connected one.

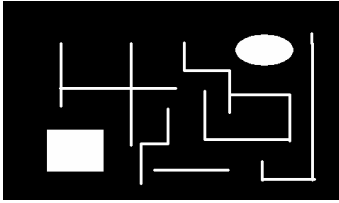


**Fig. 4.** A traditional thinning map

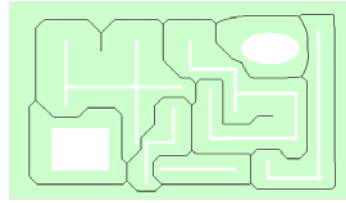


**Fig. 5.** The SCTM

Applying Algorithm 1 to a global simulation of an OGM shown in Figure 6, we get the SCTM shown in Figure 7.



**Fig. 6.** A simulation of an OGM

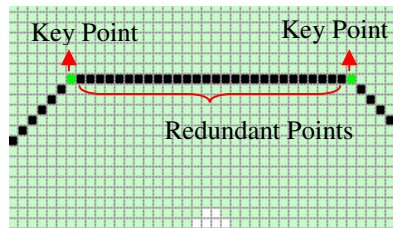


**Fig. 7.** The global SCTM

Edges of the SCTM could be used as a navigation path for mobile robots. This kind of map has many advantages. Firstly, it is a connected graph, which guarantees each node in the map is reachable, after linking any start point and any goal point to this map. Secondly, compared to the shortest path searching, although the generating path from a SCTM would be longer, we consider the path far away from surrounding obstacles plays a more important role than the shortest path, especially in complex environment. Finally, a SCTM has less branches and nodes than a Voronoi-based topological map, and has no weak meet points [6, 8], which sometimes appear in the Voronoi map and are unnecessary for navigation.

### 3 Extraction of Key Nodes

A SCTM is still a raster map even if it is composed of connected thin lines. Each thin line is composed of successive points, which has a large number and will take much time and memory to compute. In fact, a path of a robot only changes when the robot meets a junction. In other words, any two-degree point on the same line has no effects on the real path, as shown in Figure 8.

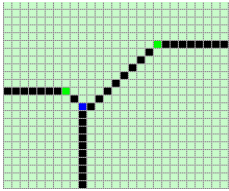


**Fig. 8.** Key points and redundant points

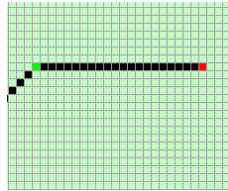
In Figure 8, a robot is certain to move from one green point to another (if the goal point is between the two green points, our algorithm will construct a path directly connecting the goal point to one of the green points), so redundant points have no effects on the representation and generation of a path. Therefore, only two endpoints on the same line are useful for the final path. We consider them key nodes. Besides, to

simplify the path searching, we ensure that each key node could reach another unique key node by moving the node in the direction of any neighbor cell of itself in the SCTM. Formally we define the concept of key nodes of a SCTM as follows:

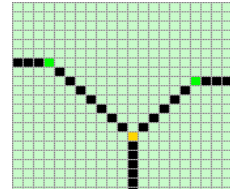
- 1) A right angle corner is a key node.
- 2) If a node  $P_1$  has no key node in its eight neighbors, then, Calculate the number of  $P_1$ 's neighbors that are '1'.
  - A. If the number is one,  $P_1$  is a key node called an end node; if it is greater than two,  $P_1$  is also a key node called a branch node.
  - B. If the number is two and  $P_1$  is the endpoint of its line, then  $P_1$  is a key node called a line node.



**Fig. 9.** A corner node (blue) and two line nodes (green)



**Fig. 10.** An end node (red) and a line node (green)



**Fig. 11.** A branch node (yellow) and two line nodes (green)

Figure 9 ~ 11 illustrate the four types of key nodes. The blue nodes denote right angle corners; red nodes denote end nodes; yellow nodes denote branch nodes; and green nodes denote line nodes. Note that eliminating the black redundant nodes does not change the connectivity of the structure of the SCTM.

We present the extraction algorithm as follows:

---

**Algorithm 2:** Extract Key Nodes

---

**Input:** A SCTM  $M_1$

**Output:** The key node graph of  $M_1$

**for** each cell  $P_1$  in SCTM

**do**

**if**  $P_2 \cdot P_1 \cdot P_4 = 1$  **or**  $P_4 \cdot P_1 \cdot P_6 = 1$  **or**  $P_6 \cdot P_1 \cdot P_8 = 1$  **or**  $P_8 \cdot P_1 \cdot P_2 = 1$   
**then** Label  $P_1$  a key node

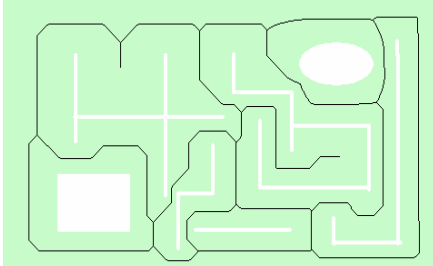
Calculate  $N(P_1)$  //  $N(P_1)$  denotes the number of neighbors of  $P_1$  that are '1'.

**if**  $N(P_1) = 1$  **or**  $N(P_1) \geq 3$   
**then** Label  $P_1$  a key node

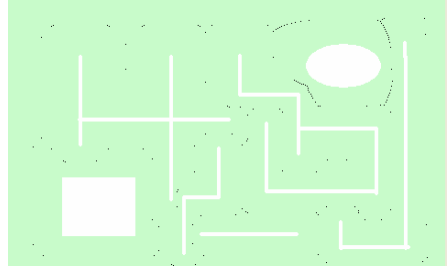
**if**  $N(P_1) = 2$  // Name the two neighbors  $n_1, n_2$   
**if**  $(P_1.Row - n_1.Row = n_2.Row - P_1.Row)$   
**and**  $P_1.Col - n_1.Col = n_2.Col - P_1.Col)$   
**or**  $n_1.Row = n_2.Row$  **or**  $n_1.Col = n_2.Col$   
**then** Label  $P_1$  a key node

---

The key node graph preserves all information of the SCTM with a smaller number of points. Applying Algorithm 2 to Figure 12, we obtain the global key node graph shown in Figure 13.



**Fig. 12.** A SCTM. Black lines denote paths



**Fig. 13.** The key node graph. Black points denote key nodes.

Extraction of key nodes makes a balance between OGM and TM. A key node graph of a SCTM holds capacity of OGM to generate a geometrical path while preserving the brevity of TM. Using a key node graph, we can easily apply the global path-searching algorithm.

## 4 Generation of Paths

For the KNG, we need a global searching strategy to generate the shortest path along edges of the graph. Our algorithm is based on A\* [9, 10], which is a kind of heuristic searching algorithm that could solve the single source shortest path problem quickly. A\* needs to calculate fitness function  $f(n)$  for each node in the graph. For each node  $n$ ,

$$f(n) = g(n) + h(n).$$

Where  $g(n)$  is the accumulated cost from start node to current node  $n$ , and  $h(n)$  is the heuristic estimated cost from  $n$  to goal node.  $h(n)$  must satisfy the condition  $h(n) \leq t(n)$  where  $t(n)$  is the actual cost from  $n$  to goal node. Every node is put in a priority queue, with one node popped each time. Each node in the map will be considered at most once.

We select the accumulated path length of the robot as  $g(n)$ , then the Euclidean distance between current node and goal node as  $h(n)$  for robot path planning problem. It satisfies the heuristic condition since the Euclidean distance is shortest between two nodes.

We still need to link the start node  $S$  and the goal node  $E$  to the KNG before applying A\* on it. We extend  $S$  and  $E$  until meeting their nearest key nodes that are denoted as  $S'$  and  $E'$ . By the same way we get the distance between each pair of adjacent nodes.

Note that Algorithm 1 and Algorithm 2 guarantee that the map is single-connected, combined with the definition of key nodes we easily draw the inference: extending any key node along the direction of its neighboring node in a SCTM. Such inference

guarantees that we could obtain and store the distance between each pair of adjacent nodes in preprocessing. Then, we present the path generation algorithm as follows:

---

**Algorithm 3:** Generate the Path

---

**Input:** A key node map  $KN_1$ ; start point S; end point E

**Output:** The shortest path along  $KN_1$  between S and E

*// preprocessing, construct the neighbor node list of  $KN_1$*

**for** each node  $k_1$  in  $KN_1$

**do**

**if** there is a neighbor  $n_1$  of  $k_1$  in the SCTM of  $KN_1$

**then** Search along  $\overline{k_1 n_1}$  until finding the unique key node  $k_2$  in this direction

Add  $k_2$  to the neighbor list of  $k_1$

*// generate the path*

Build open list and closed list, and set them empty

Link S and E to  $KN_1$ , and get S' and E'

Add S' to open list

**while** open list is not empty

**do**

currentNode = the node in open list with the lowest cost  $f(n)$

**if** currentNode = E'

**then** the path has completed

**else**

Move currentNode to closed list

**for** each node  $nc_1$  in the neighbor list of currentNode

**if**  $nc_1$  is not in open list or closed list

**then** Calculate  $f(nc_1)$

Move  $nc_1$  to open list

Show the path S, S', ..., E', E.

---

## 5 Comparison and Results

Comparison between our algorithm and traditional algorithms is listed as follows: one aspect of comparison focuses on safety and robustness.

Figure 14 shows a path in a randomly generated simulated map, with the randomly selected start node and goal node, while applying our new algorithm on it. Figure 15 shows the path generated by OGM and A\*, with the same map and nodes as Figure 14.

Figure 16 and Figure 17 is another group of figures that generated by algorithm 3 and OGM, respectively, using a part of real map generated by the laser range finder.

These figures clearly show that for our algorithm, the generated paths are not only smoother, but also safer than the traditional ones, since robots have to move along



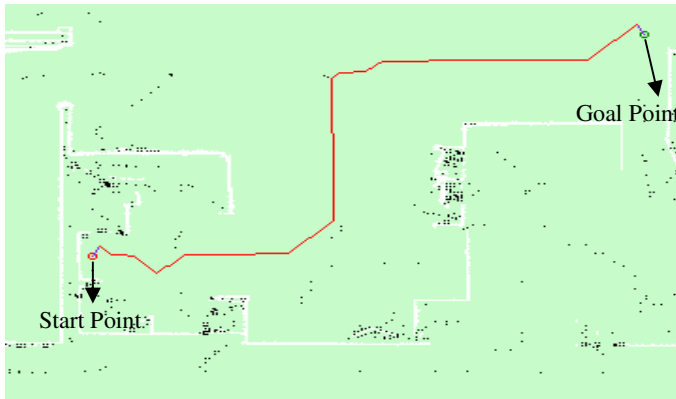


Fig. 16. A path using Algorithm 3, real map

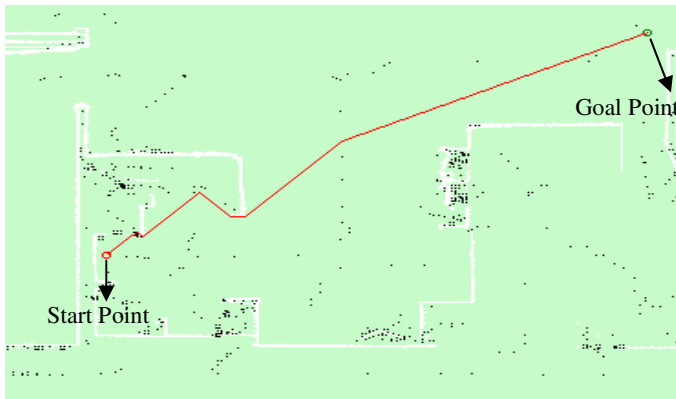


Fig. 17. A path using OGM and A\*, real map

## 6 Conclusions and Future Work

In this paper, a fast, robust robotic path planning algorithm is proposed. Concepts of SCTM and key nodes are introduced to simplify the traditional occupancy grid map. We also present effective algorithms for the generation of SCTM and the extraction of key nodes. Simulation experiments certify the effectiveness and brevity. On the basis of those algorithms we use A\* to generate a safe path quickly.

We will focus on two issues in future work. One of them is to reduce key nodes counts. We need to do interpolation for edges of irregular terrains that usually consist of dense points. The other issue is about the navigation in a wide-open space whose SCTM would be a sparse graph. In such maps, a path detaching from obstacles is not always necessary. We need to find an adaptive solution.



## Acknowledgements

The work described in this paper was partially supported by the Knowledge Innovation Program of the Chinese Academy of Sciences (Grant No.KGCX2-YW-156), the Key Laboratory of Robotics and Intelligent System, Guangdong Province (2009A060800016), and the Shenzhen Hong Kong Innovation Circle.

## References

1. Siegwart, R., Nourbakhsh, I.R.: Introduction to Autonomous Mobile Robots, pp. 185–308. MIT Press, Cambridge (2004)
2. Christoph, N., Ejan, R., Markus, G.: Genetic Path Planning for Real-time Application. In: Proceedings of the Computer Graphics International, pp. 1–8 (2004)
3. Pierre, F., Nicolas, N.: Monotony Properties of Connected Visible Graph Searching. In: Fomin, F.V. (ed.) WG 2006. LNCS, vol. 4271, pp. 229–240. Springer, Heidelberg (2006)
4. Christos, A., Paul, M.G.: Path Planning for Mobile Robot. *IEEE Transaction on System, Man, and Cybernetics* 22(2), 318–322 (1992)
5. Thrun, S.: Learning Metric-Topological Maps for Indoor Mobile Robot Navigation. *Artificial Intelligence* 98(1), 21–27
6. Kwon, T.B., Song, J.B.: Real-time Building of A Thinning-based Topological Map. *Intelligent Service Robotics* 1(3) (2008)
7. Zhang, T.Y., Suen, C.Y.: A Fast Parallel Algorithm for Thinning Digital Patterns. *Communications of the ACM* 27(3) (1984)
8. Beeson, P., Kuipers, B.: Towards Autonomous Topological Place Detection Using the Extended Voronoi Graph. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 4384–4390 (2005)
9. Fan, K.C., Lui, P.C.: Solving the Find-path Problem in Mapped Environments Using Modified A\* search Algorithm. *IEEE Transaction on System, Man, and Cybernetics* 24(9), 1390–1396 (1994)
10. Charles, W.W.: Fast Path Planning Method Using Modified A\* Method. In: IEEE International Conference on Robotics and Automation, pp. 662–667 (1993)

# Automatic Micro-manipulation Based on Visual Servoing

Xiangjin Zeng<sup>1</sup>, Yanduo Zhang<sup>1</sup>, and Xinhan Huang<sup>2</sup>

<sup>1</sup> Hubei Province Key Laboratory of Intelligent Robot,  
School of Computer Science and Engineering, Wuhan Institute of Technology,  
430073 Wuhan, China

<sup>2</sup> Dep. Of Control Science and Engineering, Huazhong University of Science and  
Technology,  
430074 Wuhan, China  
xjzeng21@sohu.com

**Abstract.** In order to perform an effective and reliable automatic microassembly, depth information estimation is the first task, a blur parameter model of defocus image based on Markov random field has been present. It converts problem of depth estimation into optimization problem. An improved Iterated Conditional Modes Algorithm has been applied to complete optimization problem, which the select of initial point employed Least squares estimate algorithm prevents that the result gets into local optimization. The visual servoing is the second task. For avoiding the complicated calibration of intrinsic parameter of camera, We apply an improved broyden's method to estimate the image jacobian matrix online, which employs chebyshev polynomial to construct a cost function to approximate the optimization value, obtaining a fast convergence for online estimation. Last, we design a PD controller to control micro-robot for completing the visual servo task. The experiments of micro-assembly of micro parts in microscopes confirm that the proposed methods are effective and feasible.

**Keywords:** Micro-assembly, Depth estimation, Broyden method, Visual servoing.

## 1 Introduction

Many simple functional MEMS devices have achieved commercial success in various applications, such as inkjet printers, accelerometers, gyroscopes, pressure sensors, displays, etc.

In order to perform an effective and reliable automatic microassembly, we must estimate firstly depth information of micro object[8][9][10][11]. As we know, the microscope vision with single CCD camera can only obtain 2D information in task space. To obtain the depth of micromanipulation from 2D plane image is a research hot topic in microscope vision.

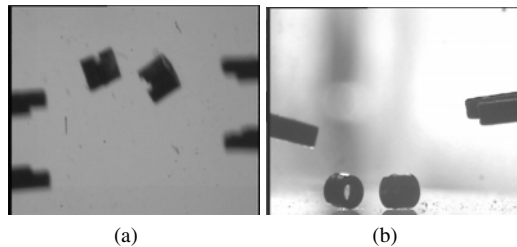
We present a blur parameter model of defocus image based on Markov random field. It converts problem of depth estimation into optimization problem. An improved Iterated Conditional Modes Algorithm has been applied to complete optimization problem, which the select of initial point employed Least squares estimate algorithm prevents that the result gets into local optimization.

Then, In order to meet the request of the high precise micro-manipulation task, robotic must employ the visual servoing method. The methods of visual servoing need calibrate precisely intrinsic parameter of camera. However, The system calibration is the complicated and difficult problem, especially for micro-manipulation based on microscope vision. So, we present the uncalibrated method to estimate image jacobian matrix online. To deal with those problems discussed above, We apply an broyden's method to estimate the image jacobian matrix. The method employs chebyshev polynomial to construct a cost function to approximate the optimization value, improving the converge of estimation.

To verify the effectiveness of the methods, Using the estimated jacobian matrix, a PD visual controller is used to make features converge to desired values with satisfactory dynamic performance. The experiments of micro-assembly of micro parts in microscopes confirm that the proposed method is effective and feasible.

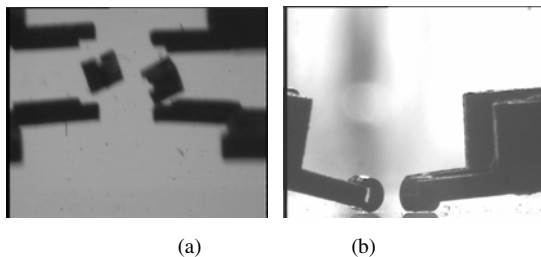
## 2 Micro-assembly Mechanisms

Micro-assembly parts size is usually  $80 \sim 1000\mu\text{m}$ , and parts's shapes include cylindrical, spherical, conical, flat-shaped, block-shaped and irregular-shaped and other types. The micro-parts assembly process is divided into clip, positioning, registration and other steps. First, we select the vacuum adsorption or the piezoelectric bimorph micro-gripper to pick up the different materials and shapes micro-objects. Second, we



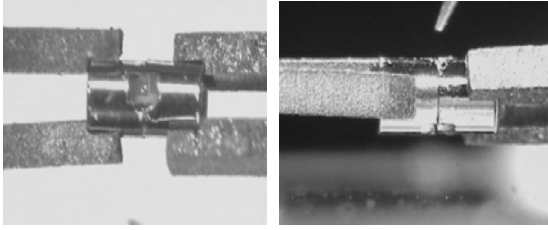
(a) Vertical view field of microscopic images (b) Horizontal view field of microscopic images

**Fig. 1.** The original microscopic image of object and the endeffector in vertical (a) and horizontal (b) view fields



(a) Vertical view field of microscopic images (b) Horizontal view field of microscopic images

**Fig. 2.** The image of endeffector automatically locating and gripping object in vertical (a) and horizontal (b) view fields



**Fig. 3.** The image of endeffector automatically locating and registration object in vertical (a) and horizontal (b) view fields

employ the corresponding micromanipulator to achieve precise positioning and assembly of positive registration. Last, the final micro-assembly process is completed.

Fig.1 shows the original optical microscopy images of un-assembly. The positioning images that micro-gripper position the center of parts is shown in Fig.2. In Fig.3, We can see that the micromanipulator assembles successfully the micro parts.

### 3 Constructing the Blur Parameter Model Based MRF

In order to operate the object in 3D space, It is a exigent problem we encountered that obtains the depth of the object. Since the change in the depth of a scene is usually gradual, the blur parameter tends to have local dependencies. Hence, we are motivated to model the blur parameter as a MRF[12]. A defocus image blur parameter model base MRF has been presented in this paper. It converts the depth problem into energy function optimization problem. Then, applies an improved Iterated Conditional Modes algorithm to optimization energy function[13], which the select of initial point employed Least squares estimate (LSE) algorithm prevents that the result gets into local optimization.

According to CCD imaging principle, we can give formula as shown in (1)

$$\frac{1}{f} = \frac{1}{u0} + \frac{1}{v0} \quad (1)$$

Where  $f$  is the focus length,  $u0$  is the distances from object to camera, and  $v0$  is the distances from image focus point to camera lens. When the distances from object to camera lens is  $u0$  and the distances from camera plane to camera lens is  $v0$ , we can obtain a clear image. If changes the distances from camera plane to camera lens, a blur image can be seen in CCD camera plane. When keeps a constant for camera parameter, It can be given the relationship between image defocus radius and the depth of image as shown in (2)

$$u = \frac{fv0}{v0 - f - Fr_b} \quad (2)$$

For formula (2),  $F$  is the F number of CCD lens.  $r_b$  is the defocus radius of image and  $u$  is the depth of the image. Therefore, There is a corresponding function relationship between the distances from CCD lens to object and the defocus radius of blur image, which can be used to obtain the depth of the image. According to formula (2), the positive or negative of distances  $u$  depends on if the focus image locates fore or back in image plane. We restrict that the distances of object is higher than the distances of image. The formula (2) can be converted into (3)

$$r_b = \frac{fv_0}{F} \left( \frac{1}{f} - \frac{1}{v_0} - \frac{1}{u} \right) \quad (3)$$

For two captured image with different focus length setting, we have formula (4)

$$r_b^i = \frac{f_i v_i}{F} \left( \frac{1}{f_i} - \frac{1}{v_i} - \frac{1}{u} \right) \quad (4)$$

For defocus image, the blur parameter is  $\rho$  and is given by  $\rho = \beta r_b$ , With giving value  $i=1, 2$  and eliminating  $u$ , the relationship of two defocus image's blur parameter is shown as (5)

$$\rho^1 = m\rho^2 + n \quad (5)$$

Where  $m$  and  $n$  are general parameter. Since the blur parameter  $\rho^i$  at location  $(x, y)$  is related to the depth of the scene, we can construct a model of the blue parameter based on MRF, meaning that the depth of the scene can be obtained indirectly.

### 3.1 MRF and Gibbs Distribution

For image function  $X$  in 2D image plane, it is thought as a 2D random field. Random variable set  $X = \{X_s : s \in S\}$ , it presumes that  $x_s$  is the realization of  $X_s$ . Pixel  $S$ 's the neighborhood is  $N_s$  and meets the conditions of probability distribution as follows:

$$P(X_s = x_s) > 0$$

$$P(X_s = x_s | X_r = x_r, r \in S, r \neq s) = P(X_s = x_s | X_r = x_r, r \in N_s)$$

We calls that  $X$  is Markov random field with neighborhood  $N_s$ . Gibbs distribution keeps a close relationship with MRF. Gibbs distribution with neighborhood  $N_s$  is expressed in formula (6)

$$P(X = x) = \frac{1}{Z} e^{-U(x)} \quad (6)$$

Where  $U(x)$  is the energy function and represents as shown in (7)

$$U(x) = -\sum_{c \in C} V_c(x) \quad (7)$$

For formula (7),  $C$  is the set of cliques included by neighborhood  $N_s$  and  $V_c(x)$  represents the potential function of clique. And  $Z = \sum_x e^{-U(x)}$  is the partition function.

### 3.2 A Blur Parameter Model Based MRF

Let  $X$  denotes the random fields corresponding to the blur parameter  $\rho^i$ ,  $X$  can be modeled by MRF. Namely, it shows as in (8)

$$P(X = x) = \frac{1}{Z} e^{-U(x)} \quad (8)$$

If  $Y_1, Y_2$  denote the random fields corresponding to the two observed images, the posterior probability can be expressed as (9)

$$P(X | Y_1, Y_2) = \frac{P(Y_1, Y_2 | X)P(X)}{P(Y_1, Y_2)} \quad (9)$$

Where  $P(Y_1, Y_2)$  is a constant and  $P(X)$  is the previous probability of the blur parameter.  $P(X | Y_1, Y_2)$  is the posterior probability of the initial image, with knowing  $Y$  value. So, according to Bayes rules, the depth restoration of the defocus image can be converted into the problem that seeks the estimation of the original image when the posterior probability is maximization. Surely, There are two main problems we have deal with. (1) computes the previous probability; (2) computes the maximization posteriori probability (MAP). Now, we give the implementation above two problems, respectively.

### 3.3 The Previous Probability Computation

Given  $X$  as the blur parameter of the defocus image, Let thinks  $X$  as a MRF, the previous probability  $P(X)$  can be used Gibbs distribution to descript.

$$P(X) = \frac{1}{Z} \left[ e^{-\sum_{c \in C} V_c(x)} \right] \quad (10)$$

For given the observed images  $y_1, y_2$ ,  $P(Y_1 = y_1, Y_2 = y_2)$  is a constant. Considering the observation model given by (11)

$$y_k(i, j) = h_k(i, j) * f(i, j) + w_k \quad (11)$$

Where  $f(i, j)$  is the clear focus image,  $y_k(i, j)$  is the blur defocus image,  $h(i, j)$  is the point spread function (PSF),  $w_k$  is the observed noise.  $h(i, j)$  has a relation corresponding to the blur radius. Following, we assume the observations of the MRF image  $y_s$  obeys the model in (12)

$$y_s = f(x_s) + w_s \tag{12}$$

Where  $f(x_s)$  is a function that maps  $x_s$  to  $\mu_{x_s}$  and  $w_s$  are independently distributed Guassian random vectors with zero mean and unknown covariance matrix  $\Theta_{x_s}$ . The PSF  $h(i, j)$  is Gaussian with blur parameter. Hence, the probability  $P(Y_1 = y1, Y_2 = y2 | X)$  can be described as Gaussian distribution and be shown as (13)

$$P(Y_1 = y1, Y_2 = y2 | X) = e^{-\sum_s \frac{1}{2\sigma^2} (y_1 - \mu_1)^2 - \sum_s \frac{1}{2\sigma^2} (y_2 - \mu_2)^2} \tag{13}$$

Then, formula (9) can be converted into (14)

$$P(X | Y_1 = y1, Y_2 = y2) = \frac{1}{Z} e^{-\sum_{c \in C} V_c(x) - \sum_s \frac{1}{2\sigma^2} (y_1 - \mu_1)^2 - \sum_s \frac{1}{2\sigma^2} (y_2 - \mu_2)^2} \tag{14}$$

Based on the observed image  $y_1, y_2$ , the problem of depth estimation is to find the estimation  $\hat{x}$  of  $X$ , which can compute the depth indirectly.

### 3.4 Improved ICM Algorithm Implementation

Base on discussion above, the posterior probability  $P(X | Y_1, Y_2)$  about the original image can be converted into the optimization problem as shown in formula (15)

$$\min \left[ -\sum_{c \in C} V_c(x) + \sum_s \frac{1}{2\sigma^2} (y_1 - \mu_1)^2 + \sum_s \frac{1}{2\sigma^2} (y_2 - \mu_2)^2 \right] \tag{15}$$

## 4 Jacobian Matrix Online Calibration

Many papers [1][3][6] have been reported some researches about image jacobian matrix online estimation. Piepmeier[4][5] presents a moving target tracking task based on the quasi-Newton optimization method. This approach is adaptive, but cannot guarantee the stability of the visual servoing. Malis's[6] method can keep the parameter of vision servo controller constant, when intrinsic parameter of camera is changed. Su J. et al[7] presents a motion 3D object tracking method based on the uncalibrated global vision feedback.

Unfortunately, the current estimation methods have problems such as estimation-lag, singularity, convergence and its speed. Especially in dynamic circumstances, these problems become more serious. To deal with those problems discussed above, We apply an broyden's method to estimate the image jacobian matrix. The method employs chebyshev polynomial to construct a cost function to approximate the optimization value, improving the converge of estimation.

### 4.1 Image Jacobian Matrix Estimation Based on Broyden Method

The image jacobian matrix can be calculated by calibrating the inner and outer parameter of robotic system & sensor system. However, it is impossible to obtain precise system parameter under a dynamic or uncertainty environment. Considering those, we employ broyden's method to estimate the image jacobian matrix.

The broyden method that solves the nonlinear equation can be shown in (16):

$$A_{k+1} = A_k + \frac{(y^{(k)} - A_k s^{(k)})s^{(k)T}}{\|s^{(k)}\|_2^2} \quad k=0,1\dots \tag{16}$$

Now, we apply the broyden method to construct estimation model of image jacobian matrix. According to equation (17), If the feature error of two images represents as

$$e(q) = f^d - f^c \tag{17}$$

Where  $f^d$  is the feature of the expectation image and  $f^c$  is the feature of the current image. The taylor series expansion of  $e_f$  is shown as

$$e_f(q) = e_f(q_m) + \frac{\partial e(q_m)}{\partial q}(q - q_m) + \dots + R_n(x) \tag{18}$$

Where  $R_n(x)$  is Lagrange remaining. We define  $J_q^*(q_n)$  as the Nth image jacobian to be estimated,

$$J_q^*(q) = \frac{\partial e(q_n)}{\partial q} \tag{19}$$

Ignoring the high order term and Lagrange remaining  $R_n(x)$ , Equation (20) can be obtained from (18) and (19), which is shown as

$$e_f(q) = e_f(q_m) + J_q^*(q_n)(q - q_m) \tag{20}$$

So, if  $J$ 's counterpoint is  $A$ ,  $\Delta e$ 's counterpoint is  $y$  and  $\Delta q$ 's counterpoint is  $s$  we can construct the image jacobian estimation model based broyden. The image jacobian estimation model based on broyden method is shown as (21).

$$J_q^*(q_{k+1}) = J_q^*(q_k) + \frac{(\Delta e - J_q^*(q_k)\Delta q)\Delta q^T}{\Delta q^T \Delta q} \tag{21}$$



The broyden algorithm estimates the optimization value by employing iterative computation. Therefore, it need the end condition of iterative computation, we employ chebyshev polynomial to construct the cost function to approximate the optimization value.

## 4.2 The Cost Function Based on Chebyshev Polynomial

Provided that

$$N_k(q) = e_f(q_k) + J_q(q)(q - q_k) \quad (22)$$

If  $N_k(q) \in c[-1, 1]$ , For chebyshev polynomial serial  $\{T_n, n = 0, 1, \dots\}$  with weight  $\rho(x) = (1-x^2)^{-\frac{1}{2}}$ , it's optimization square approximation polynomial can be shown as

$$s_n^*(x) = \frac{a_0}{2} + \sum_{i=1}^n a_i T_i(x) \quad (23)$$

where

$$a_i = \frac{2}{\pi} \int_{-1}^1 \frac{N_k(x) T_i(x)}{\sqrt{1-x^2}} dx \quad i = 0, 1, \dots, n \quad (24)$$

Then

$$N(q) = \lim_{n \rightarrow \infty} \left( \frac{a_0}{2} + \sum_{i=1}^n a_i T_i(q) \right) \quad (25)$$

Usually,  $N_k(q) \in c[a, b]$  we must convert  $c[a, b]$  into  $c[-1, 1]$ , The equation (26) can finish the convert.

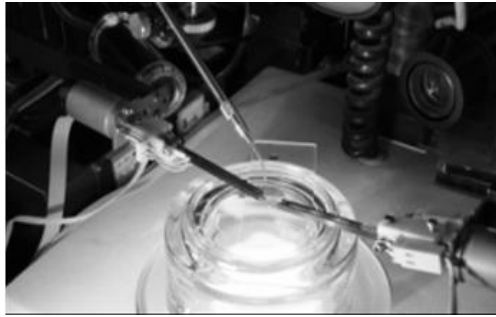
$$t = \frac{b-a}{2} x + \frac{b+a}{2} \quad (26)$$

if we use part sum  $s_n^*$  as  $N(q)$ 's approximation, under some conditions, there is a fast speed for  $a_n \rightarrow 0$

## 5 Implementation

Micromanipulation robotic system includes 3D micro-move platform, micro-gripper driven by piezoelectricity, micro-adsorption hand driven by vacuum, microscope vision and so on. Fig.4 shows the system construction of three hands cooperation micromanipulation stage.

Firstly, we construct the restoration model of the microscope vision defocus image based on MRF the same as (14). Presumes that the observed images are  $y_1, y_2$  and



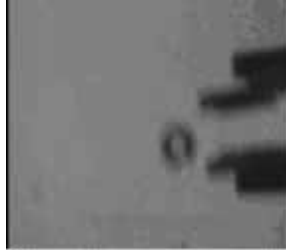
**Fig. 4.** The system construction of three hands cooperation micromanipulation stage

defines  $Y$  as the restoration image.  $Y$  is thought as a MRF. Then, we can restore the defocus image similarly as (14). During micromanipulation experiments, presumes the microscope work distances (the distances from object lens to clear imaging plane.)  $u_0 = 80\text{mm}$ , and gives that micro-move platform zero point corresponding micro-effector tip position as original point in coordinate. Then, we revise microscope vision system in order to locate original point in clear imaging plane.

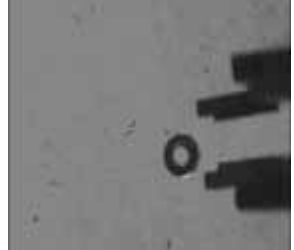
Fig.5 and Fig.6 show the initial defocus image of micro-gripper driven by piezoelectricity with different camera setting. Fig.7 gives the restoration image of micro-gripper driven by piezoelectricity.



**Fig. 5.** The initial image of defocus image micro-gripper driven by piezoelectricity

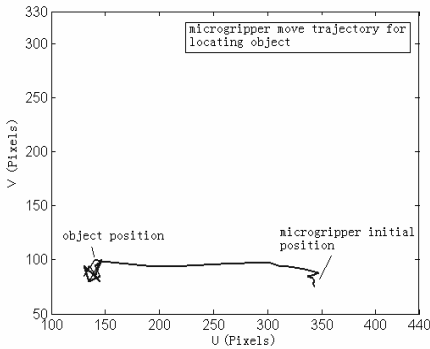


**Fig. 6.** The initial image of defocus image micro-gripper driven by piezoelectricity with different camera parameters settings

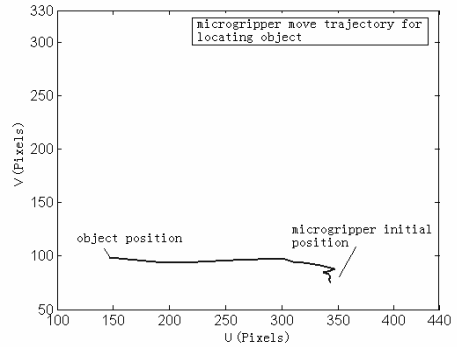


**Fig. 7.** The restoration image of defocus image of micro-gripper driven by piezoelectricity

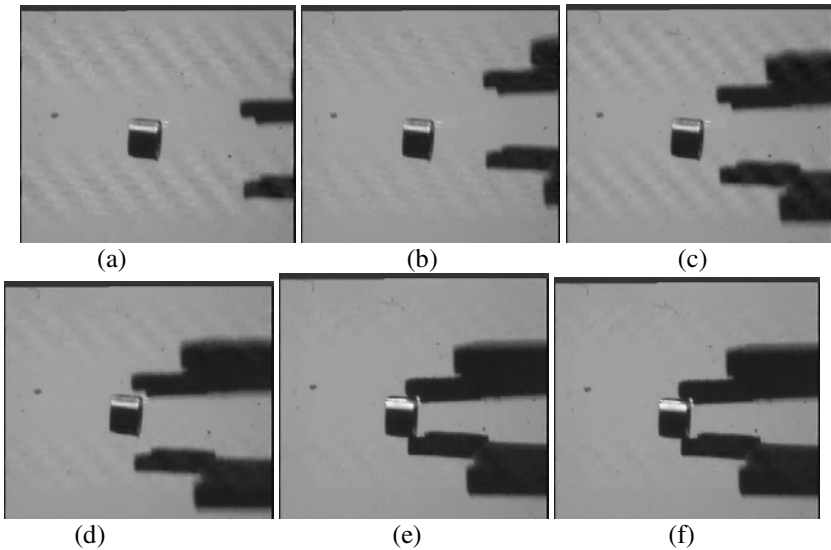
Then, we give the PD controller visual system test. Assuming that the initial parameters of PD controller  $K_p$  is 10 and  $K_d$  is 0, that is, only joined proportional control, control effect is shown in Fig.8. we can see the implementation of automatic positioning objects to the target center, a greater oscillation and overshoot. When  $K_p$  is 10 and  $K_d$  is 1.5, which incorporates proportional and differential control, control result is shown in Fig.9. Differential joined inhibits apparently the system overshoot, and the system meets the rapid and smooth.



**Fig. 8.** The trajectories of micromanipulator approaching goal objects with only proportional control (XY plane)



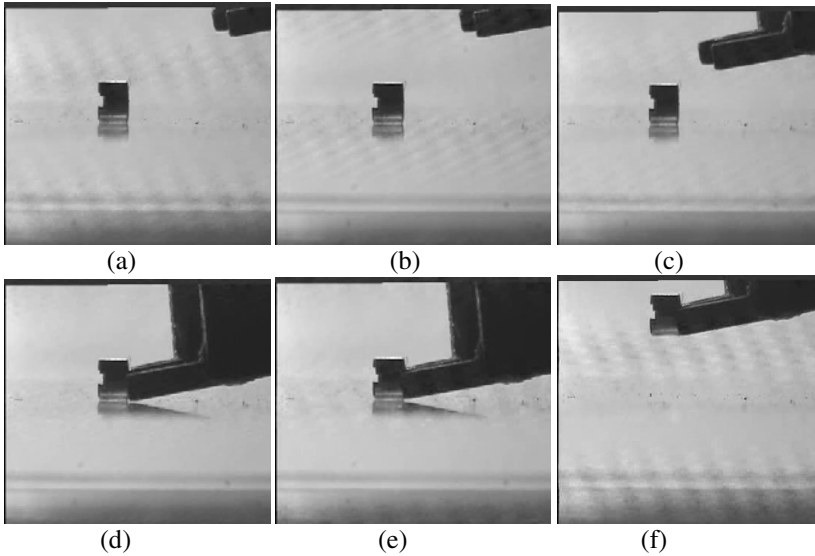
**Fig. 9.** The trajectories of micromanipulator approaching goal objects with proportional and differential control (XY plane)



**Fig. 10.** The process of the piezoelectric microgripper automatically locating and gripping the micro-target in the vertical view field

After completion of the above preparations, fully automatic microassembly tasks are performed sequentially. In order to verify the effective of uncalibrated visual servoing method, we test the experiments of single microgripper hand to position automatic and grip micro objects.

Fig.10 shows the process of the piezoelectric microgripper automatically locating and gripping the micro-target in the vertical view field. The time-consuming of process is about one minute. Fig.11 shows the process of the piezoelectric microgripper



**Fig. 11.** The process of the piezoelectric microgripper automatically locating and gripping the micro-target in the horizontal view field

automatically locating and gripping the micro-target in the horizontal view field. The time-consuming of process is about one minute:

- (a) to (c) is the process of piezoelectric microgripper close to the target micro-target;
- (d) is the process of the microgripper positioning the center of the micro target;
- (e) is the process of the piezoelectric microgripper gripping the micro target;
- (f) is the process of the piezoelectric microgripper lifting the designated height for follow-up of the micro-target assembly.

## 6 Conclusion

For the completion of three-dimensional micro-sized components assembly, a blur parameter model of defocus image based on Markov random field has been present. Then, We apply an improved broyden's method to estimate the image jacobian matrix on line, which employs chebyshev polynomial to construct a cost function to approximate the optimization value. Finally, design a PD controller to control micro-robot. In the microscopic visual environment, completes the visual servo task of micromanipulator positioning and automatic gripping small parts, the experimental results show that the algorithm is relatively satisfied with the effectiveness of implementation.

**Acknowledgments.** The project is supported by The Youth Science Fund of Wuhan Institute of Technology.

## References

1. Kang, Q.S., Hao, T., Meng, Z.D., Dai, X.Z.: Pseudo inverse Estimation of Image jacobian Matrix in Uncalibration Visual Servoing. In: Proceedings of the 2006 IEEE International Conference on Mechatronics and Automation, pp. 25–28 (2006)
2. Malik, A.S., Choi, T.S.: Consideration of illumination effects and optimization of window size for accurate calculation of depth map for 3D shape recovery. *Pattern Recognition* 40(1), 154–170 (2007)
3. Shen, S.D., Liu, Y.H., Li, K.: Asymptotic trajectory tracking of manipulators using uncalibrated visual feedback. *IEEE/ASME Trans. on Mechatronics* 8(1), 87–98 (2003)
4. Piepmeier, J.A., MacMurray, G.V., Lipkin, H.: A Dynamic Quasi-Newton Method for Uncalibrated Visual Servoing. In: IEEE International Conference on Robotics and Automation, pp. 1595–1600 (1999)
5. Piepmeier, J.A., MacMurray, G.V., Lipkin, H.: Uncalibrated Dynamic Visual Servoing. *IEEE Transactions on Robotics and Automation* 20(1), 143–147 (2004)
6. Malis, E.: Visual servoing invariant to changes in camera-intrinsic parameters. *IEEE Trans. Robot. Autom.* 20(1), 72–81 (2004)
7. Su, J., Ma, H., Qiu, W., Xi, Y.: Task-independent robotic uncalibrated hand-eye coordination based on the extended state observer. *IEEE Trans. on Systems, Man, and Cybernetics* 34(4), 1917–1922 (2004)
8. Namboodiri, V.P., Chaudhuri, S.: On defocus, diffusion and depth estimation. *Pattern Recognition Letters* 28, 311–319 (2007)
9. Tsai, D.M., Lin, C.T.: A moment preserving approach for depth from defocus. *Pattern Recognition* 31(5), 551–560 (1998)
10. Ziou, D., Deschenes, F.: Depth from defocus estimation in spatial domain. In: *Computer Vision and Image Understanding*, pp. 143–165 (2001)
11. Deschenes, F., Ziou, D., Fuchs, P.: An unified approach for a simultaneous and cooperative estimation of defocus blur and spatial shifts. In: *Image and Vision Computing*, pp. 35–57 (2004)
12. Rajagopalan, A.N., Chaudhuri, S.: An MRF model based approach to simultaneous recovery of depth and restoration from defocused images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(7), 577–589 (1999)
13. Park, J., Kurz, L.: Image enhancement using the modified ICM method. *IEEE Transactions on Image Processing* 5(5), 765–771 (1996)
14. Malik, A.S., Choi, T.S.: Consideration of illumination effects and optimization of window size for accurate calculation of depth map for 3D shape recovery. *Pattern Recognition* 40, 154–170 (2007)

# Self-calibration of a Stewart Parallel Robot with a Laserranger\*

Yu Liu<sup>1,2</sup>, Hong Liu<sup>2</sup>, Fenglei Ni<sup>2</sup>, Wenfu Xu<sup>1</sup>, and Feng Han<sup>2</sup>

<sup>1</sup> Academy of Fundamental and Inter-disciplinary Sciences, Harbin Institute of Technology,  
Harbin 150001, China

<sup>2</sup> State Key Laboratory of Robotics and Systems, Harbin Institute of Technology,  
Harbin 150001, China

**Abstract.** In order to overcome influence of kinematic parameter errors on end-effector pose (position and orientation) accuracy of the Stewart parallel robot, a new self-calibration method with a laserranger fixed on the mobile platform is presented. Firstly, a new calibration scheme is proposed and the direct kinematic model of the parallel robot is built. Then the distance error model is derived, and the objective function of calibration is given. Subsequently, an improved differential evolution algorithm (DE) is adopted to identify the kinematic parameters. At last, simulation is made in a set of validation positions to verify the above calibration method. The results show that after calibration pose accuracy of the parallel robot has a significant improvement. At the same time, the strong stochastic search ability of DE is validated.

**Keywords:** Stewart parallel robot; laserranger; error model; differential evolution.

## 1 Introduction

Parallel robots are characterized by low inertia, high stiffness, great load-to-self-weight ratio and better repetition, and widely used in such many occasions as setting a broken bone for medical treatment, simulating ocean wave for shipbuilding and machining. In view of these tasks, tracking predefined trajectories precisely is the basic function of parallel robots. However better repetition does not mean higher positioning accuracy, subject to machining and assembly error, the real kinematic parameters of a parallel robot generally deviate from their nominal ones, further which will cause the end-effector pose (position and orientation) errors. Since calibration is a cost-effective method to improve pose accuracy of a parallel robot, it is necessary to calibrate it before it is applied to kinds of occasions.

Many literatures developed kinds of calibration methods of parallel manipulators, mainly they are categorized into two classes, namely classical calibration [1] [2] and self-calibration [3] [4] [5]. Using classical calibration methods the real transformation matrix from the base coordinate frame to the mobile platform coordinate frame can

---

\* Foundation item: Project (60775049 and 60805033) supported by National Natural Science Foundation of China.

usually be obtained by an outer measurement device. Besides, its nominal transformation matrix can be evaluated according to its direct kinematics with its nominal kinematic parameters. The difference between the two transformation matrixes can be used as the optimized objective function of the parallel robot calibration. However, because the inverse kinematics of the parallel robot has usually the analytical solution and is more easily evaluated, more classical calibration methods are represented as the form of solving inverse kinematics, namely the objective function is denoted by the measurement residual which is the difference between the measured and computed joint variable. For the classical calibration methods, their external measurement devices mainly include theodolite [1], CMM [2], OPTOTRAK [6], vision [7], laser displacement sensors [8] and so on.

At present self-calibration develops fast and wins much advance. Usually, self-calibration is also divided into two types, one is to exert outer physical constraints on the parallel robot, and the other is to install the extra sensors on it. For the former, Khalil and Besnard [9] fixed a passive universal joint or a passive spherical joint using a lock mechanism and presented a fully autonomous calibration method for a six DOF parallel robots. Abdul and Jeha [10] locked the mobile platform at a specific point with a device and made it lose 3 translational degrees of freedom to calibrate a parallel Hexa Slide manipulator. However, the 1<sup>st</sup> type self-calibration method needs the backdrivability of the actuators, which is difficult to accept for the actuators with a large reduction ratio [11]. Comparatively, the 2<sup>nd</sup> type self-calibration method acquires more attention. Adding 3 redundant sensors between the forearms and the base, Ecorchard and Maurine [12] proposed a self-calibration method with elastic deformation compensation for Delta Parallel Robots, and the experiment showed the calibration led to a seven time improvement of the capability. Zhuang [4] installed two additional sensors on each of the alternative U-joints of a Stewart platform, and by creating a suitable measurement residual using conflicting information provided with redundant sensing, self-calibration of the parallel robot was realized. Besides, by means of angular passive joint sensors [13], the self-calibration approaches were studied. In this paper, a laserranger fixed on the mobile platform of a parallel robot is used to calibrate its kinematic parameters, and its details are discussed later.

Still, we know that many calibration models were solved based on Levenberg-Marquardt method [3] [10] [13]. However, subject to disturbance of measurement noise, the calibration algorithm might be unstable. So, some calibration methods based on evolution algorithm arises [14] [15], the paper adopts DE to identify the kinematics parameters of the 6-DOF Stewart platform.

## 2 Description of Error Model

### 2.1 Direct Kinematics

As shown in figure 1, the Stewart parallel robot consists of a fixed base and a mobile platform connected with six legs of drivable and variable length. Each leg is linked to the mobile platform by the S-joint and to the base by the U-joint. Let  $A_i$  and  $B_i$  respectively denotes the centre of the  $i^{\text{th}}$  S-joint and U-joint and  $l_i$  does the length  $A_i B_i$  of the  $i^{\text{th}}$  leg ( $i = 1, \dots, 6$ ). At the same time we define the tool

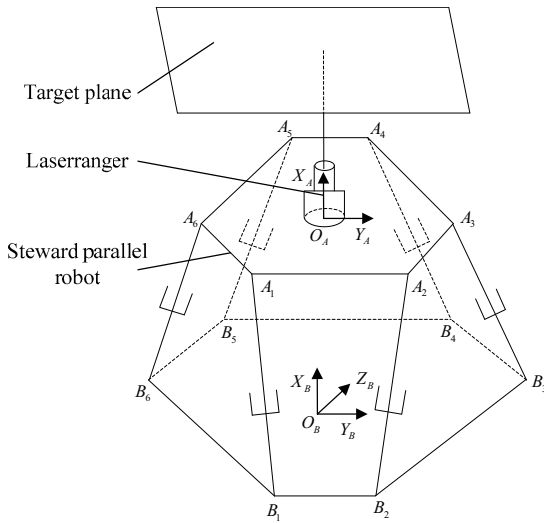


Fig. 1. Calibration scheme of Stewart parallel robot

coordinate frame  $A\{X_A, Y_A, Z_A\}$  with respect to the platform and the fixed coordinate frame  $B\{X_B, Y_B, Z_B\}$  with respect to the base. Suppose that the homogeneous coordinate of the point  $A_i$  relative to the frame  $A$  is  ${}^A\mathbf{a}_i(a_{ix}, a_{iy}, a_{iz}, 1)$ , similarly, that of the point  $B_i$  relative to the frame  $B$  is  ${}^B\mathbf{b}_i(b_{ix}, b_{iy}, b_{iz}, 1)$ , and the transformation matrix from the frame  $B$  to the frame  $A$  is  ${}^B T_A$  (given by the external measurement device). Accordingly the homogeneous coordinate  ${}^B\mathbf{a}_i$  of the point  $A_i$  relative to the frame  $B$  can be written as

$${}^B\mathbf{a}_i = {}^B T_A {}^A\mathbf{a}_i \tag{1}$$

Since the transformation matrix  ${}^B T_A$  can be denoted by a pose vector  $X = [x, y, z, \alpha, \beta, \gamma]$ ,  ${}^B\mathbf{a}_i$  is the function of the six independent parameters of the vector, where  $\alpha, \beta$  and  $\gamma$  denotes z-y-x Euler angle. Namely

$${}^B\mathbf{a}_i = {}^B\mathbf{a}_i(x, y, z, \alpha, \beta, \gamma) \tag{2}$$

When the six driven displacements  $l_i (i=1, \dots, 6)$  are given, the direct kinematic model of the Stewart platform can be built.

$$l_i^2 = ({}^B\mathbf{a}_i - {}^B\mathbf{b}_i)^T ({}^B\mathbf{a}_i - {}^B\mathbf{b}_i) \quad (i=1, \dots, 6) \tag{3}$$

The above formula denotes a nonlinear equation set. With Eq. (3), it is difficult to obtain analytical solution of the direct kinematics, and up to 40 solutions may exist



[16]. A numerical iterative method based on the conventional Newton-Raphson method is used to find a local direct solution. The procedure can be summarized as

- 1) Set an initial pose vector  $\mathbf{X}_0$  ;
- 2) Let

$$f_i = l_i^2 - \left( {}^B\mathbf{a}_i - {}^B\mathbf{b}_i \right)^T \left( {}^B\mathbf{a}_i - {}^B\mathbf{b}_i \right) \quad (i = 1, \dots, 6) \tag{4}$$

- 3) According to the following equation evaluate  $\Delta\mathbf{X}$

$$\mathbf{f}^{(m)} + \left[ \frac{\partial \mathbf{f}}{\partial \mathbf{X}} \right]^{(m)} \Delta\mathbf{X} = \mathbf{0} \tag{5}$$

where  $\mathbf{f} = [f_1 \ \dots \ f_6]^T$  ,  $\Delta\mathbf{X} = [\Delta x \ \Delta y \ \Delta z \ \Delta\alpha \ \Delta\beta \ \Delta\gamma]^T$  ,  $m$  indicates the  $m^{\text{th}}$  iteration.

- 4) Calculate  $\mathbf{X}_0 = \mathbf{X}_0 + \Delta\mathbf{X}$  ;
- 5) If  $\|\mathbf{f}\|_{\infty} > \varepsilon$  , then go to (3), else end.

So,  $\mathbf{X}$  can be evaluated according to the known  $l_i$  , further  ${}^B T_A$  will be done.

### 2.2 Distance Equation

According to Fig. 1, it is easily seen that a laserranger is fixed on the mobile platform, and there is a target plane in the front of it, so the laserranger can measure the distance between the mobile platform and the settled target. Because the tool frame  $A$  can be chosen arbitrarily, we might as well let the laserranger coordinate frame act as the tool coordinate frame, namely, the starting point  $\mathbf{p}_s$  of the laser beam is located in the origin  $O_A$  and the laser beam consists with the positive direction of the axis  $X_A$  . Besides, the matrix  ${}^B T_A$  can be divided into the following sub-matrix.

$${}^B T_A = \begin{pmatrix} {}^B \mathbf{R}_A & {}^B \mathbf{p}_A \\ 0 & 1 \end{pmatrix} \tag{6}$$

where  ${}^B \mathbf{R}_A \in R^{3 \times 3}$  denotes an Orientation matrix,  ${}^B \mathbf{p}_A \in R^3$  does a translational vector. According to the Eq. (6) and the above definition of the laserranger coordinate frame, it is easy to know that  $\mathbf{p}_s$  relative to the frame  $B$  equals to the translational vector  ${}^B \mathbf{p}_A$  .

Similarly, the laser beam unit vector  $\mathbf{b}_i$  with respect to the coordinate frame  $B$  is represented as

$$\mathbf{b}_i = {}^B \mathbf{R}_A \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \tag{7}$$

Suppose that the measured plane equation in the coordinate frame  $B$  is

$$\mathbf{n}_l \mathbf{p} + f = 0 \tag{8}$$

where  $\mathbf{n}_l(n_{lx}, n_{ly}, n_{lz})$  is the normal vector of the target plane, its positive direction can be chosen arbitrarily, here  $n_{lz}$  is assigned a positive value.  $\mathbf{p}$  denotes the coordinate vector  $(p_x, p_y, p_z)$  of the arbitrary point in the target plane, and  $f$  is a known scalar. If the laser beam vector  $\mathbf{b}_l$  intersects the target plane at the point  $\mathbf{p}_j$ , obviously in terms of the position relation of the vectors,  $\mathbf{p}_j$  can be written as

$$\mathbf{p}_j = \mathbf{p}_s + h\mathbf{b}_l \tag{9}$$

where  $h$  denotes the distance from the point  $\mathbf{p}_s$  to the intersecting point  $\mathbf{p}_j$ . We know,  $\mathbf{p}_j$  also meets the Eq. (8), by substituting Eq. (9) into (8)  $h$  can be represented as follows.

$$h = -\frac{\mathbf{n}_l \mathbf{p}_s + f}{\mathbf{n}_l \mathbf{b}_l} \tag{10}$$

### 2.3 Error Model

The calculated distance  $h$  from Eq.(10) is an estimate value based on the nominal  $s$  of the parallel robot. In fact, because of the errors resulted from manufacturing, assembly and other factors the real coordinate  ${}^A\mathbf{a}_i^r$  and  ${}^B\mathbf{b}_i^r$  usually deviate from their nominal values  ${}^A\mathbf{a}_i$  and  ${}^B\mathbf{b}_i$ , besides,  $l_i$  will also have an initial offset. Since 3 parameters for each point need to be calibrated, again adding 6 link length parameters the identified parameters amount to 42.

Therefore, the distance error  $\Delta h$  caused by the geometrical parameter errors can be written as

$$\begin{aligned} \Delta h &= h({}^A\mathbf{a}_1^r, \dots, {}^A\mathbf{a}_6^r, {}^B\mathbf{b}_1^r, \dots, {}^B\mathbf{b}_6^r, l_1^r, \dots, l_6^r) - h({}^A\mathbf{a}_1, \dots, {}^A\mathbf{a}_6, {}^B\mathbf{b}_1, \dots, {}^B\mathbf{b}_6, l_1, \dots, l_6) \\ &= h^r - h \end{aligned} \tag{11}$$

where  $l_i^r$  denotes the real leg length,  $h^r$  is the real distance. In terms of the Eq. (11), the optimized objective function of the parallel robot self-calibration can be represented as

$$\min H = \sqrt{\sum_{j=1}^m \Delta h_j^2} \tag{12}$$

where  $H$  is the optimized objective function,  $m$  denotes the number of calibration positions, and it is much larger than the number of the identified parameters.

### 3 Geometrical Parameter Identification Based on DE

#### 3.1 Differential Evolution Algorithm

Similar to the genetic algorithm (GA), DE is a kind of global evolutionary algorithm, presented by R. Storn and K. Price in 1995. DE has proven to be one of the best evolutionary algorithms.

DE is encoded by real number, whose initial populations are composed of  $Np$ ,  $D$ -dimensional real-valued parameter vectors. DE has three operators, i.e. mutation, crossover and selection, which is the same as GA. However, their order is contrary. Among them mutation is the most original, in which the new parameter vectors are created by adding the weighed difference between the two population vectors chosen randomly to the third vector, namely

$$\mathbf{x}'_i = \mathbf{x}^t_a + F(\mathbf{x}^t_b - \mathbf{x}^t_c) \quad (13)$$

where,  $\mathbf{x}'_i$  denotes the  $i^{\text{th}}$  mutated vector,  $\mathbf{x}^t_a, \mathbf{x}^t_b, \mathbf{x}^t_c$  represent the different parameter vectors chosen randomly in the  $t^{\text{th}}$  generation population, and  $F$  is the scaling factor of the difference. Here,  $i$  differs from  $a, b, c$ .

In fact, the conventional DE mutation still include another strategy, namely

$$\mathbf{x}'_i = \mathbf{x}^t_{Best} + F(\mathbf{x}^t_b - \mathbf{x}^t_c) \quad (14)$$

where  $\mathbf{x}^t_{Best}$  is the best parameter vector in the population. The strategy is generally called as DE/best/1, while the former is done as DE/rand/1. However, it is found in the simulation that DE/rand/1 excessively lays emphasis on the diversity of the created individual, so its convergence is inferior and the desired optimization level is difficult to be attained. On the other hand, DE/best/1 lets the best parameter vector in the population become the search centre, which limits the diversity of the created individual and initially leads to fast optimization but later easily gets into local minima. In view of this, a new tradeoff strategy is proposed in the paper, namely

$$\mathbf{x}'_i = (\mathbf{x}^t_a + \mathbf{x}^t_{Best}) / 2 + F(\mathbf{x}^t_b - \mathbf{x}^t_c) \quad (15)$$

With the above formula, optimization capability of DE has a greater improvement.

After mutation of each parameter vector, the non-uniform crossover with a crossover constant  $CR \in [0,1]$  between the mutated vector  $\mathbf{x}'_i$  and target vector  $\mathbf{x}_i$  is used to create the trial vector  $\mathbf{v}_i$ . Its  $j^{\text{th}}$  component  $v_{ij}$  is represented as follows

$$v_{ij} = \begin{cases} \mathbf{x}'_{ij}, & \text{rand}(j) \leq CR \\ \mathbf{x}_{ij}, & \text{rand}(j) > CR \end{cases} \quad (16)$$

where  $\text{rand}(j)$  is a uniformly distributed random number in the range of  $[0,1]$ . The above equation shows that  $CR$  is a crossover probability, which controls the components of the trial vector provided with the mutated vector and the target vector. If  $CR$

is larger, the mutated vector occupies the trial vector more. When  $CR$  becomes 1, the trial vector  $v_i$  equals to  $x_i'$ . DE is more sensitive to the choice of  $F$  than that of  $CR$ , while  $CR$  is more like a fine tuning element. The higher  $CR$  is, the faster convergence is.

At last, selection is made for reproduction. DE uses a tournament mechanism by which the trial vector competes against its target vector. If the objective function value of the trial vector is lower than that of the target vector, it will survive to the next generation, vice versa. The selection function is given as

$$x_i(t+1) = \begin{cases} v_i(t), & f(v_i(t)) \leq f(x_i(t)) \\ x_i(t), & f(v_i(t)) > f(x_i(t)) \end{cases} \quad (17)$$

### 3.2 Parameter Identification Process

Subsequently, DE is used for the kinematic parameter identification of the Stewart platform. As stated previously, DE has few parameters ( $F$  and  $CR$ ), and is easy to program and solve the complicated optimization problem. The parameter vector  $x$  attached to the population is determined as follows.

$$x = [\Delta a_{1x} \Delta a_{1y} \Delta a_{1z} \cdots \Delta a_{6x} \Delta a_{6y} \Delta a_{6z} \Delta b_{1x} \Delta b_{1y} \Delta b_{1z} \cdots \Delta b_{6x} \Delta b_{6y} \Delta b_{6z} \Delta l_1 \cdots \Delta l_6] \quad (18)$$

where each element of the vector  $x$  is the error value between the real and nominal geometrical parameter.

Then the calibration procedure is listed as follows.

1. Give initial values to DE parameters, i.e.  $N_p$ ,  $CR$ ,  $F$ , the maximum iteration number  $G$  and the stopping criterion  $V$  of the optimization. Here  $F$  and  $CR$  are both generally in the range  $[0.5, 1]$ , the number of parents  $N_p$  is set 5 to 10 times the number of parameters,  $G$  and  $V$  are determined according to the actual case.
2. Calculate the objective function values of the parameter vectors, if meet  $H \leq V$ , go to 9;
3. Give the constrained scopes of the identified parameters;
4. Set  $i=1$ , and initialize the population ( $N_p \times D$ ) with the stochastic selection method in the constrained scopes;
5. Mutate each parameter vector of the current population with the Eq.(10) by randomly choosing the three different embers in the population;
6. Create the crossover trial vectors with the Eq.(18);
7. Make the trial vector compete against its target vector by comparing their objective function values, and the winning parameter vector will succeed to the next generation;
8.  $i = i + 1$ , if  $i < G$ , go to 2;
9. End the calibration and obtain the optimal parameter vector.

## 4 Simulation of Parallel Robot Calibration

### 4.1 Consideration of Measurement Noise

$h^r$  in Eq. (11) can be given through the laserranger in this example. However there are some inevitable measurement errors in the measured distance values, which will create the disadvantageous impacts on Stewart parallel robot calibration. In order to simulate the real case maximally, measurement noise should be added to the error model (11) so as to make the calibration results more believable and effective. Here, it is assumed that distance measurement noise of the laserranger follows a normal distribution with zero mean and standard deviation 0.2mm.

In order to reduce disturbance of stochastic measurement noise, for each calibration position the distance measurement will be repeated many times. After the mean of these measurement values are evaluated, it is added to the calibration model as the measurement result in the calibration position. Of course, it is not enough only through average filter. On the other hand, more redundant calibration positions are used for the parallel robot calibration to minimize effect of measurement noise.

The distance measurement value  $h^c$  from the laserranger can be written as

$$h^c = h^r + \xi \quad (19)$$

where  $\xi$  denotes measurement noise. Then the distance error  $\Delta h^c$  between the measurement value and the calculated one with the nominal parameters can be given as follows.

$$\Delta h^c = h^c - h = \Delta h + \xi \quad (20)$$

So, subject to measurement noise, the objective function (12) can be modified as

$$H = \sqrt{\sum_{j=1}^m (\Delta h_j + \xi)^2} \quad (21)$$

where  $m$  denotes the number of the calibration positions.

### 4.2 Simulation Results

Subsequently, through simulation calibration of the parallel robot is implemented to verify the above algorithm. Totally, 150 typical calibration positions are chosen. The simulation includes two cases, namely calibration without noise filter and calibration with average filter. The parameter value of DE is designated as follows.

$$N_p = 210, CR = 0.9, F = 0.45, G = 250$$

Fig.1 (a) and (b) respectively represent optimization process of the objective function corresponding to the first and second case. For the average filter, repeated measurement times in each calibration position are chosen as 10. It is easy to see that in the beginning the optimization curves decline steeply, and then turn gentle little by little, at last become flat, which shows that DE behaves very stable and possesses strong stochastic search ability in spite of disturbance of larger measurement noise.

Table 1 gives the nominal kinematic parameters and real errors of the Stewart parallel robot and the identified parameter errors corresponding to the above two cases. Table 2 gives a comparison of pose errors after robot calibration in a set of independent validation positions. In nature, robot calibration is a fit for measurement data in the calibration positions, so it is necessary to search for extra positions (namely validation positions) to verify calibration effect. A discount of pose accuracy is inevitable in the validation positions compared with calibration positions. Here the position and orientation errors all indicate the resultant errors of the three components of the position or orientation vectors, for instance, the resultant position error is represented as  $\sqrt{x^2+y^2+z^2}$ . Attentively, in principle orientation errors can not be synthesized because of mutual dependence of its three components, but here they are a tiny magnitude and can be considered independent each other. As shown in Table 2, for the first case, the maximum position and orientation errors respectively decrease from 2.8689mm and 2.1329mrad to 0.7726mm and 0.3748mrad. In contrast, for the second case these errors decrease to 0.2376mm and 0.1672mrad. Similarly, comparison of the RMS (root mean square) pose errors can be given. It is easy to see that pose accuracy of the parallel robot has a significant improvement after calibration. Moreover, the calibration results for the second case are better than that for the first case, which shows importance of the filter.

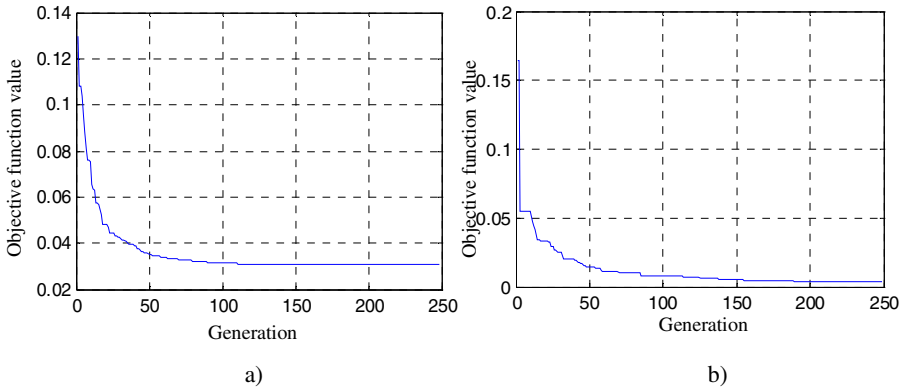


Fig. 2. Variation of the objective function values

Table 1. Identified kinematic parameter errors in two cases

Parameter	Nominal Value (mm)	Real Errors (mm)	Identified Errors without filter (mm)	Identified Errors with filter (mm)
$a_{1x}$	3	0.0297	-0.2943	0.3327
$a_{1y}$	254	0.3382	0.1951	0.2966
$a_{1z}$	52	0.3976	0.4557	0.4710
$a_{2x}$	2	0.1463	0.5995	0.1971
$a_{2y}$	165	0.3249	-0.1831	0.2938
$a_{2z}$	205	0.6402	0.9254	0.5868

**Table 1.** (continued)

$a_{3x}$	-1	-0.3358	0.2136	-0.4707
$a_{3y}$	-81	-0.1838	-0.3206	-0.1642
$a_{3z}$	-246	-0.3929	0.7797	-0.2574
$a_{4x}$	2	0.8195	0.1588	0.9957
$a_{4y}$	101	0.6132	-0.5126	0.6865
$a_{4z}$	-243	-0.2179	-0.5798	-0.3545
$a_{5x}$	-2	-0.3099	-0.5704	-0.4309
$a_{5y}$	-172	-0.1944	0.5457	-0.1283
$a_{5z}$	193	0.4512	-0.1640	0.3354
$a_{6x}$	0	-0.2714	-0.3652	-0.2581
$a_{6y}$	-253	-0.9486	0.2959	-0.8920
$a_{6z}$	35	0.7716	0.9971	0.6710
$b_{1x}$	-7	-0.9641	-0.9980	-0.8434
$b_{1y}$	-9	-0.7384	-0.6162	-0.7110
$b_{1z}$	92	0.1041	0.3013	0.1621
$b_{2x}$	39	0.7098	-0.0789	0.6448
$b_{2y}$	161	0.4867	0.1936	0.3810
$b_{2z}$	108	0.5594	0.1970	0.4586
$b_{3x}$	-11	-0.4112	-0.2406	-0.5042
$b_{3y}$	82	0.3695	-0.2595	0.4563
$b_{3z}$	-34	-0.7277	0.9593	-0.6613
$b_{4x}$	41	0.1028	0.0841	0.1502
$b_{4y}$	11	0.8711	-0.9470	0.7607
$b_{4z}$	-190	-0.2989	0.0478	-0.3319
$b_{5x}$	-12	-0.1908	0.0341	-0.2022
$b_{5y}$	-72	-0.9932	0.2417	-0.8766
$b_{5z}$	-49	-0.9537	-0.8419	-0.9697
$b_{6x}$	41	0.0711	-0.6022	0.0317
$b_{6y}$	-172	-0.4813	-0.1188	-0.5863
$b_{6z}$	88	0.4867	1.0012	0.5443
$l_1$	561.96	0.5	0.6993	0.6105
$l_2$	550.63	0.5	-0.5798	0.4317
$l_3$	562.62	0.5	0.6056	0.5994
$l_4$	553.94	0.5	0.3102	0.4090
$l_5$	559.99	0.5	0.7513	0.3798
$l_6$	554.51	0.5	-0.2216	0.4780

**Table 2.** Pose error comparison after robot calibration

Error Type	RMS Position Error (mm)	RMS Orientation Error (°)	Maximum Position Error (mm)	Maximum Orientation Error (°)
Nominal parameter	2.3746	1.3746	2.8689	2.1329
Identified parameter (no filter)	0.6359	0.2142	0.7726	0.3748
Identified parameter (with filter)	0.1268	0.0978	0.2376	0.1672

## 4 Conclusion

In the paper, a new self-calibration method with a laserranger is adopted to improve pose accuracy of the Stewart parallel robot. It has the following advantages.

1. The calibration method does not rely on external measurement systems to measure the end-effector pose, moreover the whole calibration system is simple, so the technique is much easier and more convenient to implement.
2. By changing the original mutation operator, the differential evolution algorithm is improved on. Improved DE both lays emphasis on calculative efficiency and given attention to optimization level. In spite of more kinematic parameters, DE still implements the calibration task of the Stewart parallel robot effectively, which proves its strong stochastic search ability.
3. Simulation results show that the average filter can effectively decrease the bad influence of measurement noise on calibration. Besides, increase of calibration positions can also restrain it.

However, it is pointed out that such non-geometrical errors as joint and link flexibility, backlash etc. are not considered in the paper. If higher pose accuracy is required, these factors are no longer negligible and must be paid attention to.

## References

1. Zhuang, H., Masory, O., Yan, J.: Kinematic Calibration of Stewart Platforms Using Pose Measurements Obtained by a Single Theodolite. In: Proceedings of the IROS, pp. 329–335. IEEE Press, New York (1995)
2. Cong, D.C., Yu, D.Y., Han, J.W.: Kinematic Calibration of Parallel Robots Using CMM. In: Proceedings of the 6th World Congress on Intelligent Control and Automation, pp. 8514–8518. IEEE Press, New York (2006)
3. Zhuang, H., Liu, L.: Self Calibration of a Class of Parallel Manipulators. In: Proceedings of the IEEE Int. Conf. on Robotics and Automation, pp. 994–999. IEEE Press, New York (1996)



4. Zhuang, H.: Self Calibration of Parallel Mechanisms with a Case Study on Stewart Platforms. *IEEE Transaction on Robotics and Automation* 13, 387–397 (1997)
5. Phillip, L., Christoph, B., Jürgen, H.: Self-Calibration of the HEXA-Parallel-Structure. In: *Proceedings of the IEEE Int. Conf. on Automation Science and Engineering*, pp. 393–398. IEEE Press, New York (2005)
6. Bai, S.P., Ming, Y.T.: Kinematic Calibration and Pose Measurement of a Medical Parallel Manipulator. In: *Proceedings of the IEEE Int. Conf. on Control Automation Robotica and Vision*, pp. 419–424. IEEE Press, New York (2002)
7. Renaud, P., Andreff, N., et al.: Vision-Based Kinematic Calibration of a H4 Parallel Mechanism. In: *Proceedings of the IEEE Int. Conf. on Robotics and Automation*, pp. 1191–1196. IEEE Press, New York (2003)
8. Maurine, P., Dombre, E.: A Calibration Procedure for the Parallel Robot Delta 4. In: *Proceedings of the IEEE Int. Conf. Robotics and Automation*, pp. 975–980. IEEE Press, New York (1996)
9. Khalil, W., Besnard, S.: Self Calibration of Stewart–Gough Parallel Robots without Extra Sensors. *IEEE Transaction on Robotics* 15(6), 1116–1121 (1999)
10. Abdul, R., Jeha, R.: Fully Autonomous Calibration of Parallel Manipulators by Imposing Position Constraint. In: *Proceedings of the IEEE Int. Conf. Robotics and Automation*, pp. 2389–2394. IEEE Press, New York (2001)
11. Abdul, R., Aslam, P., Jeha, R.: Experimental Results on Kinematic Calibration of Parallel Manipulators using a Partial Pose Measurement Device. *IEEE Transaction on Robotics* 22(2), 379–384 (2006)
12. Ecorchard, G., Maurine, P.: Self-Calibration of Delta Parallel Robots with Elastic Deformation Compensation. In: *Proceedings of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Edmonton, pp. 1283–1288. IEEE Press, New York (2005)
13. Last, P., Budde, C.: HEXA-Parallel-Structure Calibration by Means of Angular Passive Jjoint Sensors. In: *Proceedings of the IEEE Int. Conf. on Mechatronics and Automation*, pp. 1300–1305. IEEE Press, New York (2005)
14. Storn, R., Price, K.V.: Minimizing the Real Functions of the ICEC 1996 Contest by Differential Evolution. In: *Proceedings of the IEEE Conference on Evolutionary Computation*, pp. 842–844. IEEE Press, New York (1996)
15. Xu, Q.S., Li, Y.M.: Error Analysis and Optimal Design of a Class of Translational Parallel Kinematic Machine Using Particle Swarm Optimization. *Robotica* 27, 67–78 (2009)
16. Husty, M.L.: An Algorithm for Solving the Direct Kinematic of Stewart-Gough-Type Platforms. *Mechanism and Machine Theory* 31(4), 365–380 (1996)

# Real-Time Binaural Sound Source Localization Using Sparse Coding and SOM

Do-hyeong Hwang and Jong-suk Choi

Korea Institute Science and Technology (KIST),  
Cheongryang, Seoul 130-650, Korea  
{rapperdo, cjs}@kist.re.kr

**Abstract.** This paper presents a binaural sound source localization method using both a sparse coding and a self-organizing map (SOM) in real-time system. We use the sparse coding for feature extraction to estimate the azimuth of sound source. It is used for decomposing input sound signals into three components such as time, frequency and magnitude. Therefore, although the frequency characteristic of ITD (Interaural Time Difference) is changed by shape of head, we utilized it to estimate the azimuth of the sound source considering the time-frequency features simultaneously. Then we adapted the SOM to estimate the azimuth of sound source which is a type of artificial neural networks. This system is constructed by open-source software, Flowdesigner, which gives us a data-flow oriented developmental environment for efficient real-time system.

**Keywords:** real-time binaural sound source localization, sparse coding, SOM, flowdesigner.

## 1 Introduction

As robotics research and industry have been growing, many kinds of robots have been developed for interaction with humans in various places. During the development, the shape of robots was similar to that of humans for more friendly interaction. Also, there have been many research studies about sound source localization system because sound source localization is one of the interactions between human and robot.

As a general rule, time delay of arrival (TDOA) or interaural time difference (ITD) is used in the sound source localization system as a cue to estimate azimuth of sound source. However, it is difficult to apply this method to humanoid robot because the method assumes a free field between microphones. In case of, interaural time difference (ITD) is largely changed depending on the frequency of the sound, because of the shape of the head bone, even though the sound source is generated from the same spot. To overcome this problem, we implemented new sound source localization system using sparse coding and SOM [11]. Sparse coding can decompose a sound signal into spikes which have three components: time, frequency and magnitude. Then we can acquire the ITD feature from the spikes, and self-organizing map (SOM) is applied to the results of the sparse coding to estimate the azimuth. After performing neuron's learning in the SOM, we can estimate the azimuth of the sound source through the SOM. We implemented all of system in real-time using a Flowdesigner

framework which is data-flow oriented development environment. It can be used to build complex applications by combining small and reusable building blocks. Therefore we made several Flowdesigner modules to complete real-time binaural sound source localization system.

## 2 Hardware Configuration

Our robot is consisted of two ears, head and body as shown in Fig 1. We used KE-MAR ear and the head was manufactured in Korea Institute of Science and Technology (KIST). An 8-channel analog to digital converter (TD-BD-8CSUSB, Tokyo Electron Device Ltd.) is installed in the robot. The body of this robot is PeopleBot made by Active Media Robotics.



Fig. 1. The robot that has a head, two ears and body

## 3 Software Configuration

### 3.1 Flowdesigner as Framework of the System

In order to implement the efficient real-time system, we used a Flowdesigner [10] framework and implemented several component blocks. Then, we built a binaural sound source localization system, which can guarantee the real-time operation. We developed some modules such as SparseCoding, Mean Variance Filter, Coefficient Filter, SOMLocalization and so on.

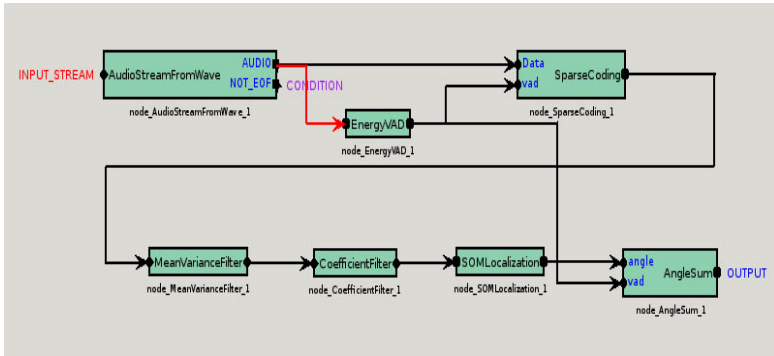


Fig. 2. Configuration of the real-time sound source localization system in Flowdesigner

### 3.2 The Flow of System Configuration and Real-Time Evaluation

Our system is consisted of some modules such as Fig. 2. We acquired the 2-ch sound signal through the DAQ board, and each input signal is transformed into spikes by the Sparse Coding module. After getting the spikes, we calculated the S-ITD from binaural spike (we will explain the meaning of S-ITD in the next chapter). Also, we used three kinds of filters to get the clear S-ITD. Finally, we estimated the azimuth of sound source by filtered S-ITD and trained SOM Localization module. We calculated the RTF (Real-Time Factor) value to evaluate the real-time ability. The real time factor is defined as Eq. (1).

$$RTF = \frac{P}{I} \tag{1}$$

Where  $P$  is processing time and  $I$  is an input duration. Data structure of our system is composed of 3200 samples and is shifted while 25% of them are overlapped. One frame is 150ms because our sampling frequency is 16kHz. Therefore, the total processing time is 145.0ms and RTF value is 0.97. Our system is operated in real-time since the RTF is less than 1. Table 1 shows the processing time of each module.

Table 1. Processing time of each module

Module name	Average of processing time(msec)
AudioStreamFromMic	62.53
SparseCoding	82.4
MeanVarianceFilter and CoefficientFilter	0.043
SOM Localization	0.03
Total Time	145.01

## 4 Feature Extraction Using Sparse Coding

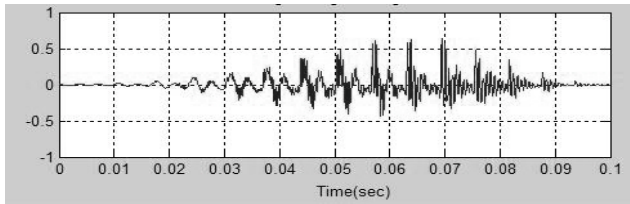
### 4.1 Sparse and Kernel Representation

In order to acquire the robust ITD at different frequency range, we adapted encoding technique. We used a sparse and shiftable Kernel method of sound signal representation [1] since this method can decompose a sound signal into three components: time, frequency and magnitude. In this method, the sound signal  $x(t)$  is encoded with a set of kernel functions that can be positioned arbitrarily and independently in time. The sound signal will be expressed as Eq. (2) with ‘‘Spikes’’ which have their own kernel number, varying amplitude and position.

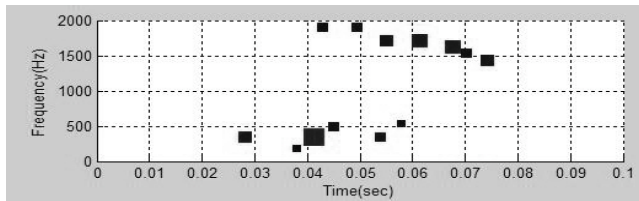
Here the kernel function we select is a Gammatone filterbank whose center frequency and width are set according to an ERB (Equivalent Rectangular Bandwidth) filterbank cochlear model [2] so as to obtain the spike from different frequency range. Eventually we can obtain the ITD from different frequency range. We set 64 as the number of filterbank.

$$x(t) = \sum_{m=1}^M \sum_{i=1}^{n_m} s_i^m \phi_m(t - \tau_i^m) + \varepsilon(t) \quad (2)$$

where,  $\tau_i^m$  is temporal position,  
 $S_i^m$  is coefficient of the  $i^{\text{th}}$  instance of kernel  $\phi_m$ ,  
 $n_m$  indicate the number of instances of  $\phi_m$ ,  
 $M$  is the number of  $\phi_m$ ,  
 $\varepsilon(t)$  is a noise.



(a) Original sound signal



(b) Decomposed signal by sparse coding

**Fig. 3.** A Sound Signal and spikegram

Fig.3 illustrates a generative model, which represents a sound signal as a set of kernel functions; Fig. 3 (a) is the original sound signal and (b) is the representation of the signal as a set of kernel functions. We call this plot a spikegram, and call squares the spikes in various sizes located in a spikegram. In the spikegram, x-axis is time, y-axis is frequency (center frequency of gammatone filterbanks), and the size of square is the magnitude of spike. In order to extract the spikes from the original sound, matching pursuit algorithm [3] is used for producing a more efficient estimate of the time positions and coefficients.

### 4.2 Interaural Time Difference

As mentioned in the introduction, we used the ITD from the spike as a cue of binaural sound localization. In order to obtain the spike from each microphone signal, we performed the Matching pursuit algorithm. In this algorithm, the signal can be decomposed into Eq. (3), where  $\langle x(t), \phi_m \rangle$  is inner product between the signal and the kernel and  $R(t)$  is the residual signal. More generally, we can express the equation as Eq. (4), where n is number of iteration with  $R_x^0(t) = x(t)$  at the start of the algorithm. Essentially we can obtain the set of spikes referring to  $S_i^m(m, \tau_i^m)$ .

$$x(t) = \langle x(t), \phi_m \rangle \phi_m + R(t) \tag{3}$$

$$R_x^n(t) = \langle R_x^n(t), \phi_m \rangle \phi_m + R_x^{n+1}(t) \tag{4}$$

$$\phi_m = \arg \max_m \langle R_x^n(t), \phi_m \rangle \tag{5}$$

Then we calculated the time difference of two spike sets which is called S-ITD (Sparse coding based ITD).

However, it is difficult to find a pair of spike sets from two channel microphone signal including time delay information. In order to find the pairs of spike, we adapted a modified matching pursuit algorithm after first channel was decomposed. In modified matching pursuit algorithm, the kernel m is selected by previous decomposed bank number of spike and we limited a time index range of  $\tau_i^m$ ,  $\tau_{i_s}^m - ND < \tau_i^m < \tau_{i_s}^m + ND$ , where ND is a number of sample index maximally delayed and  $\tau_{i_s}^m$  is a position of previously decomposed spike in other channel.

The time difference has errors that cause failure in the estimation of the direction of sound source. The errors can be removed easily using characteristic of spike. The first thing to solve the problem is mean-variance filter. The mean can be calculated by the summation of Gaussian, which has the S-ITD time positions as the mean. The second is bandpass filter. We use a spike of which the bank of center frequency is from 500Hz to 4,000Hz because it is frequency range of voice. The last filter is threshold filter in the magnitude (coefficient) of the spike, i.e., the time difference that has smaller coefficient is filtered out. Using these three kinds of filters, we can remove undesirable spikes and get the precise S-ITD.

## 5 Classification by Self-organizing Map

### 5.1 Sound Collection

In our research, we fabricate a robot system that collects sound at -70 degrees to +70 degrees with 5 degree resolution, and build up a database for the learning of SOM.

### 5.2 Learning

In order to estimate the direction of sound sources from the extracted features, we used the Self-Organizing Map (SOM) [4] which is a type of neural network. We can easily expand 3D sound localization and expect a good estimation performance since SOM uses a neighborhood function to preserve the topological properties of the input space. It also utilized competitive learning as training data so we can expect to implement an on-line learning easily.

We organized 1-dimension SOM at each frequency bin because S-ITD has a different characteristic in frequencies. The SOM had 58 neurons for the representation of azimuth angle and 3,000 iterations for the organizing map. The Fig.4 shows the results of learning neurons at some banks. In this figure, x-axis is index of trained neurons and y-axis is their weight. It represents the characteristics of each S-ITD. The learning algorithm is demonstrated in steps.

Step 1. Initialize weight vectors and neighborhood size.

Step 2. Select a 'winning node'. The winning node is selected by Euclidean distance between input and weight vector. This equation is like this,

$$C(t) = \arg \min_i (\|X(t) - W_{ij}(t)\|_2) \quad (6)$$

Where,  $C(t)$  is the winning node

$W(t)$  is the weight vector of node  $i$  at time  $t$ .

$X(t)$  is the input vector.

Step 3. Update the weights using equation (7) ~ (9)

$$W_{ij}(t+1) = W_{ij}(t) + \Delta W_{ij}(t). \quad (7)$$

$$\Delta W_{ij}(t) = \alpha \phi(i, k) (X - W_{ij}(t)) \quad (8)$$

$$\phi(i, k) = e^{\frac{-d(i, c)^2}{\beta(t)^2}} \quad (9)$$

Where  $\phi(i, k)$  is referred to as the neighborhood size,

$\alpha$  is a learning rate,

$-d(i, c)^2$  is the Euclidean distance from node  $i$  to the winning node  $c$ ,

$\beta(t)^2$  is the neighborhood size at time  $t$ .

## 6 Evaluation

We evaluated the system performance at each 29 difference angles using recorded sound of voice. The recorded sentences consisted of “silbot come here” for training and “hello, silbot welcome” for evaluation of performance in the same time period 5 second. The distance between the robot and a speaker is one meter and the SNR (Signal to Noise Ratio) is 12~15dB. Fig. 5 shows statistical graph of SD and RMSE at each angle. The red points represent the mean value and the green rectangles mean a standard deviation at each angle. The average of total standard deviation is 2.09 degree and the average of total RMSE is 2.55 degree. The results show that using the S-ITD feature by the sparse coding is possible to accurately estimate the azimuth angle for sources.

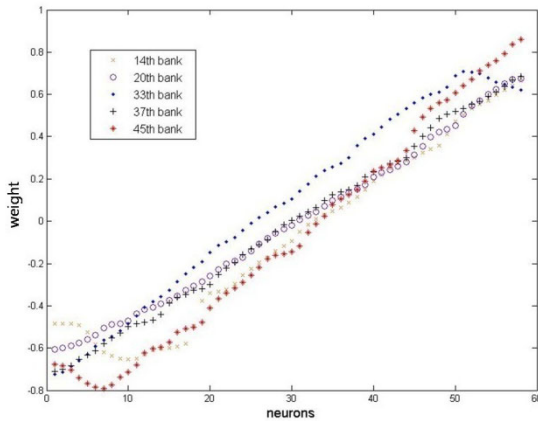


Fig. 4. The results of learning at neurons of some banks

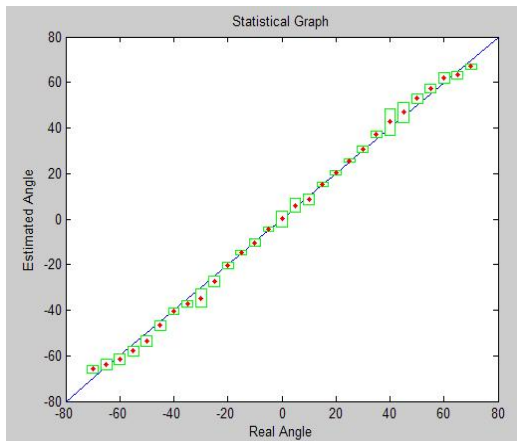


Fig. 5. Statistical graph of 1m experiment



## 7 Conclusion

In this paper, we have proposed a new binaural sound localization system with humanoid robot which is implemented in real-time environment based on the sparse coding and SOM.

After making the spikes of each ear's signals, calculating the S-ITD between the binaural spikes, and applying three kinds of filtering of the S-ITD, we have got clean S-ITD data. Also, we could estimate azimuth angle of sound sources with high performance by using learning method of SOM. Moreover, we implemented the whole system which can be guaranteed of the real-time operation.

This is a new and innovative approach on sound source localization with both sparse coding and SOM. Our next plan is to implement a 3D sound source localization using this characteristic of spike and to apply various environments using online learning of SOM.

## References

1. Lewicki, M.S., Sejnowski, T.J.: Coding time-varying signals using sparse, shift-invariant representations. *Advances in Neural Information Processing Systems* 11, 730–736 (1999)
2. Slaney, M.: Auditory toolbox. Technical Report 1998-010, Interval Research Corporation
3. Mallat, S.G., Zhang, Z.: Matching Pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing* 41(12), 3339–3415 (1993)
4. Kohonen, T.: The self-organizing map. *Proceedings of the IEEE* 78(9), 1464–1480 (1990)
5. Performance PeopleBot Operations manual
6. Nakashima, H., Mukai, T.: 3D Sound source localization system based on learning of binaural hearing. In: *III International Conference on Systems, Man and Cybernetics*, vol. 4, pp. 3534–3539 (2005)
7. Smith, E.C., Lewicki, M.S.: Efficient coding of time-relative structure using spikes. *Neural Comput* 17, 19–45 (2005)
8. Calmens, L., Lakemeyer, G., Wagner, H.: Azimuthal sound localization using coincidence of timing across frequency on robotic platform. *J. Acoust. Soc. Am.* 121(4), 2034–2048 (2007)
9. Keyrouz, F., Diepold, K.: An Enhanced Binaural 3D Sound Localization Algorithm. In: *isspit*, pp. 662–665 (2009); *IEEE International Symposium on Signal Processing and Information Technology* (2006)
10. Flowdesigner is a free “data flow” development environment which get, <http://sourceforge.net/projects/flowdesigner/>
11. Kim, H.-s., Choi, J.-s.: Binaural Sound Localization based on Sparse Coding and SOM. In: *Proceedings of the 2009 IEEE/RSJ international conference on Intelligent Robots and Systems*, pp. 2557–2562 (2009)

# Design and Control of a Hydraulic-Actuated Leg Exoskeleton for Load-Carrying Augmentation

Heng Cao<sup>1</sup>, Jun Zhu<sup>1</sup>, Chunming Xia<sup>1</sup>, Hong Zhou<sup>2</sup>, Xiao Chen<sup>2</sup>, and Yu Wang<sup>1</sup>

<sup>1</sup> School of Mechanical and Power Engineering,  
East China University of Science and Technology  
200237 Shanghai, China

<sup>2</sup> The quartermaster research institute of general logistics department of PLA  
100082 Beijing, China  
hengcao@ecust.edu.cn

**Abstract.** For the research on control method of leg exoskeleton for carrying payload, a leg exoskeleton prototype called ELEBOT (ECUST Leg Exoskeleton roBOT) has been built. To increase the load-carrying ability of exoskeleton, pseudo-anthropomorphic mechanical structure, a kind of efficient, low-flow pump-valve hybrid controlled hydraulic actuator, custom sensing shoes and other electrical components for control have been developed. A knee torque estimate algorithm based on GRF (ground reaction force) was applied to ELEBOT via a hydraulic direct force feedback PD controller. Experiments indicated ELEBOT could support itself and 30 kg additional payload with above control method when it was walking at a speed up to 3.2 km/h while following the operator's maneuvers. These results indicate the control method discussed in this work is valid.

**Keywords:** exoskeleton, hydraulic, servo, force feedback.

## 1 Introduction

An exoskeleton is a powered mechanical device that is “worn” by an operator and work in concert with the operator's movements. Most researchers of exoskeleton focused on two main applications. One type of exoskeleton is developed for torque and work augmentation, and the other type is developed for load transfer [1].

The first type could improve walking and running metabolic economy or might be used to reduce joint pain or increase joint strength in paralyzed or weak joints. HAL (hybrid assistive leg) has been developed by Prof. Yoshikuyi and his research team [2]. HAL used EMG sensors and floor reaction force sensors to obtain the conditions of exoskeleton and the operator.

The second type provides a “bypass” to transfer the load weight directly to the ground so that it could benefit the operator in load carrying by increasing load capacity, lessen the likelihood of leg or back injury, and improve metabolic locomotory economy. The most successful example of this kind is BLEEX (Berkeley lower extremity

exoskeleton) developed by Human Engineering Laboratory of UC Berkeley [3]. Users wearing BLEEX could support a 75 kg load while walking at 0.9 m/s.

A leg exoskeleton prototype called ELEBOT has been developed for load-carrying augmentation. ELEBOT is simpler and more energy saving than full-actuated leg exoskeletons such as BLEEX since only one DOF is actuated: knee flexion/extension. An efficient, low-flow pump-valve hybrid controlled hydraulic actuator has been developed for farther reducing the energy consumption. Direct force feedback PD controller has also been used to reduce the impedance of hydraulic actuator. The experiments indicated ELEBOT could efficiently assist operator in level-ground walking with 30 kg payload. The performance specification of ELEBOT is shown in Table 1.

The design details of mechanical structure, hydraulic actuator, sensing and control system, as well as the experiment results of load-carrying walking will be discussed in following sections.

## 2 Hardware of ELEBOT

For the goal of load-carrying augmenting, mechanical structure of exoskeleton should be lightweight, highly elastic and easy to fit different operators, the actuator should be powerful, compliance and energy-saving, and the sensing and control system should be simple and reliable.

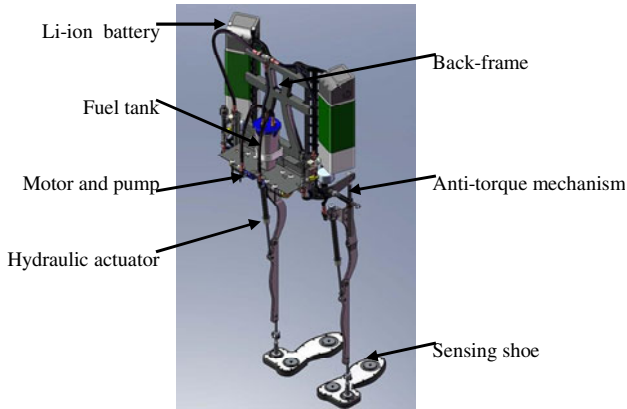
**Table 1.** Performance Specification of ELEBOT

Subject	Performance Specification
Weight	Total weight without batteries: 22 kg Battery: two 4.5 kg batteries
Power	Li-ion Batteries 80-volt system
Electronics	Custom on-board MCU controller; Signal Conditioning Circuit
Hydraulics	Standard hydraulic fluid Efficient low-flow, pump-valve hybrid controlled hydraulic actuator Normal operation at 100 bar, peak operation up to 140bar
Operator height	Between 1.60 m and 1.90 m
Payload	up to 30 kg
Max speed	3.2 km/h with 30 kg payload 4.5 km/h without payload

### 2.1 Mechanical Structure

The mechanical structure of ELEBOT is pseudo-anthropomorphic, which means the leg is kinematically similar to a human's, but does not include all of the degrees of freedom of human legs, and ELEBOT is easier to fit various operators.

Pseudo-anthropomorphic lower extremity of ELEBOT has seven distinct degrees of freedom per leg: 3 degrees at the hip, 1 degree at the knee (pure rotation in the sagittal plane) and 3 degrees at the ankle. The human hip is a ball and socket joint with three degrees of freedom [4]. Though knee is a complex joint, a pure rotary in the sagittal is



**Fig. 1.** The mechanical structure of ELEBOT with hydraulic actuators and batteries

chosen for simplicity and robustness. The ankle joint has three degrees of freedom just like the human's ankle. The joint ranges of motion are determined by examining human joint ranges of motion so that its kinematics is close to human's. All unactuated joints are spring-loaded and the lengths of thigh and shank are both adjustable to fit operators between 1.60 m and 1.90 m. The only actuated joint of ELEBOT is knee joint, since a study has indicated that torque increasing at knee is much higher than that at ankle and hip when a human is walking with a payload [5]. The knee joint also constitutes a key factor for stability and for ensuring a smooth and natural movement [6].

The back-frame is used to mount hydraulic components, control system and payload. It is also one of rigid connections at the extremities of leg (another connection is foot). In order to reduce the effect of torque due to the exoskeleton and payload applied to operator, all components and payloads are mounted close to the back of operator.

An anti-torque mechanism is also designed at hips for further improving. This mechanism is a rotating guide-bar mechanism which is made up of belt, thigh and a gas spring. With a constraint on thigh, anti-torque mechanism is functional only when the relevant leg is in stance phase.

## 2.2 Hydraulic Actuator

Hydraulic actuators have higher specific power than EM motors, thus they are one kind of ideal actuator of exoskeleton, particularly for load-carrying augmentation. An efficient, low-flow pump-valve combined hydraulic actuator is designed for ELEBOT, which mainly consists of a permanent magnet torque motor directly driving a small displacement internal gear pump, a high-speed On-Off valve and a bidirectional linear hydraulic cylinder.

Traditional valve-controlled electro-hydraulic servo system has some drawbacks, such as complex structure, high energy consumption, large heat emission, high precision requirement of oil filtrating and big pressure losing, particularly heavy weight and

expensive. Compared with aforesaid system, the pump-controlled electro-hydraulic servo system has property of compact structure, better reliability, and higher mechanic efficiency, particularly the good performance of saving energy [7]. However, the dynamic response of pure pump-controlled system is lower than valve-controlled system. For this reason, an unexpressive high-speed On-Off valve is used to improve the performance of hydraulic actuator and to avoid additional complexity for motor reverse.

Our previous work [8] indicates that the performance of exoskeleton system will be better when exoskeleton has low-impedance. Hydraulic actuator has high impedance that is not suitable to a man-machine system like exoskeleton. A direct force feedback controller is used to reduce the impedance of hydraulic actuator. This means any contact forces due to the position difference between exoskeleton and operator will be minimized [9]. An accumulator is used to stabilize the pressure of pump output when the pump is working and to save the hydraulic power when the high-speed On-Off valve is active.

Control problems such as dead band and friction can be reduced significantly in the pulse-width modulation (PWM) control manner. The schematic diagram of a hydraulic force servo system is showed in Fig. 2.

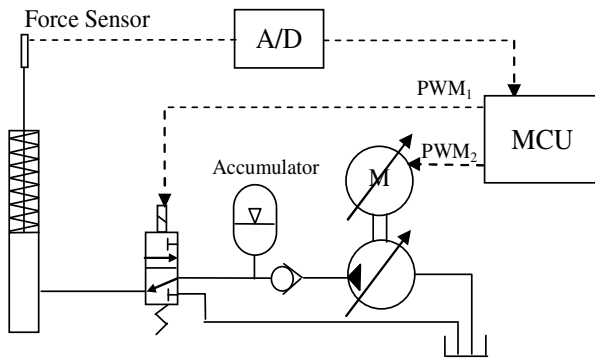


Fig. 2. The diagram of hydraulic actuator

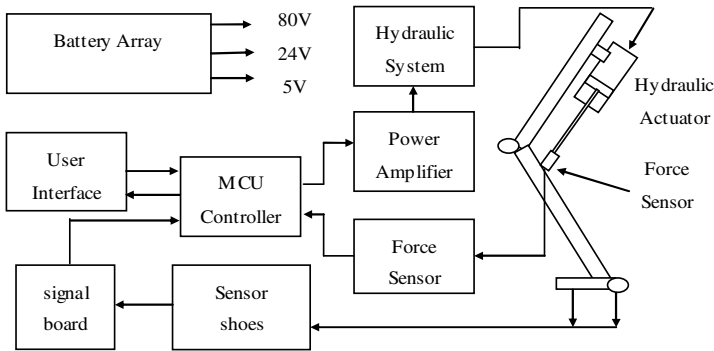
### 2.3 Sensing and Control System

Two kinds of sensor interfaces are used in ELEBOT: sensor shoes and force sensor located at the end of the hydraulic piston rod.

In order to detect the important parameters and features of gait, three force sensors (spoke type pressure transducer) are installed in the sole of each sensing shoe (shown in Fig. 3). Sensor A is close to the toe, sensor B is located at the heel and sensor C is under the extremity of the exoskeleton leg. With the contacting signal of each shoe provided from sensor A and B, the controller chooses the correct control model to be applied. Sensor A and B also provide the GRF acts on operator. Lastly, Sensor C can measure actual load transfer to the ground via the exoskeleton.



**Fig. 3.** The sensing shoe of ELEBOT. A, B and C are the force sensor in the shoe.



**Fig. 4.** The hardware structure of sensing and control system

The force sensor located at the end of the hydraulic piston rod can acquire the actual output of actuator only when actuator is static since the additional torque  $\tau_c$  that operator act on exoskeleton will also be acquired while the exoskeleton is moving. Further process is necessary to extract useful force signal information by using an effective algorithm.

The signal preamplifier circuit is the critical link of signal extraction, which is to amplify weak signals of sensors (mV) and to filter noises. Since the output signal of sensors is weak and includes large common mode voltage, it is sensitive to the external disturbance and suffers from null shift, noise and instability. Chopper-stabilized amplifier is used in the signal preamplifier circuit design and zeroing circuit ensures the output signal of signal preamplifier circuit is normalized.

The controller of ELEBOT is developed based on a MCU, and the control system of exoskeleton is real-time, robust, high-precision and parameter-adjustable.

For the moment, ELEBOT is powered by two common heavy Li-ion batteries. The weight of each battery is about 4.5 kg. Selection for a new lighter and efficient power source is underway. A customized power circuit is used to provide the electrical apparatus of exoskeleton with difference voltages.

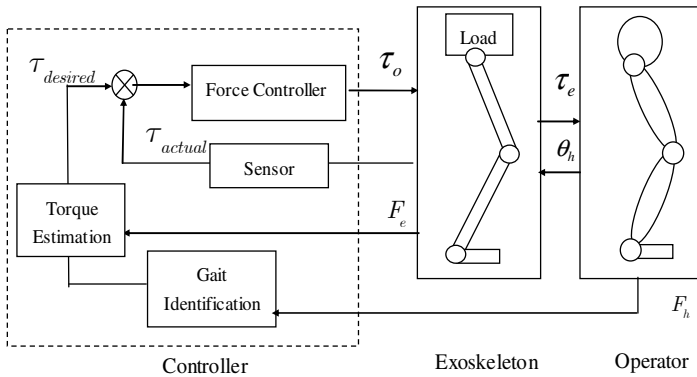


Fig. 5. The signal flow and command flow of ELEBOT

### 3 Control of ELEBOT

The purpose of control system is to guarantee the exoskeleton can mimic operator's gait and transfer the payload to the ground synchronously [10]. As a device operates in parallel with the human legs, the exoskeleton should provide appropriate torque and transfer the payload to the ground. We think the gait of operator is the most basic command for exoskeleton to detect the operator's intent. Furthermore, since the GRF is the only external force acting on a walking person except the gravity, which means GRF is exclusively responsible for accelerating the center of mass of human during locomotion [11]. A simple algorithm based on GRF is used to control exoskeleton via controlling GRF. In brief, the control of ELEBOT consists of three parts: gait identification, knee torque estimation based on GRF and force feedback control.

#### 3.1 Gait Identification

A gait cycle is defined as initiating with the heel strike of one foot and ending at the next heel strike of the same side. According to the state of GRF, the gait cycle may be subdivided into two phases: stance phase and swing phase. The gait information can be detected by sensing shoes, as well as GRF acts on operator. During stance phase, the weight of the exoskeleton including payload is transferred entirely to the ground through the stance leg. Since the swing leg is barely loaded, the knee joint is not controlled for energy-saving and simpleness of control. The swing leg can move with the operator naturally.

#### 3.2 Knee Torque Estimate Algorithm Based on GRF

In order to keep the acceleration of exoskeleton with that of operator, the proportion between GRF acts on exoskeleton and on operator should be closed to the proportion between the mass of exoskeleton and that of operator.

If factor  $\lambda$  is a desired carrying capacity of exoskeleton, it means that,

$$\lambda = m_e / (m_h + m_e). \tag{1}$$

Where  $m_e$  is the mass of exoskeleton,  $m_h$  is the mass of operator and  $\alpha$  is the acceleration of human. Then the desired GRF acts on exoskeleton,  $F_d$ , can be estimated as:

$$F_d = \lambda (F_e + F_h). \tag{2}$$

Where  $F_e$  is GRF acts on exoskeleton, it can be measured by sensor C in sensing shoe.  $F_h$  is GRF acts on operator, which can be measure by sensor A and B.

And the desired knee torque  $\tau_d$  also can be estimated as:

$$\tau_d = R \times F_d \tag{3}$$

If we assume that  $F_d$  is pointed to the hip joint, than  $R$  is a function of the knee angle  $\theta$ , the length of thigh  $R_1$  and the length of shank  $R_2$ .

$$R = \frac{R_1 R_2 \sin(\theta)}{\sqrt{R_1^2 + R_2^2 - 2R_1 R_2 \cos(\theta)}}. \tag{4}$$

### 3.3 Force Feedback Controller

A direct force feedback controller is implemented to control the hydraulic actuator servo to the desire force,  $F_d$ , which is calculated by knee torque algorithm. It is also benefit to eliminate any contact forces due to the position difference between exoskeleton and operator for reducing the impedance of hydraulic actuator. A PD controller is used for this force servo control.

$$u(t) = \begin{cases} 0 & |e(t)| \leq E_1 \\ K_p e(t) + K_d \frac{de(t)}{dt} & E_1 \leq |e(t)| \leq E \\ 1 & E_2 \leq |e(t)| \end{cases} \tag{5}$$

Where  $e(t)$  is the error between  $F_d$  and actual force feedback by force sensor.  $u(t)$  is output width of PWM,  $K_p$  is the proportionality coefficient,  $K_d$  is the derivative coefficient. If  $e(t) > 0$ , controller will increase speed of motor so that the pressure of hydraulic actuator will be increased. If  $e(t) < 0$ , controller will activate the high-speed On-Off valve to decrease the pressure of hydraulic actuator. To avoid vibration of control, the output signals are clipped when  $|e(t)|$  is out of the rang between the prescribed thresholds  $E_1$  and  $E_2$ .

## 4 Experiment and Discussing

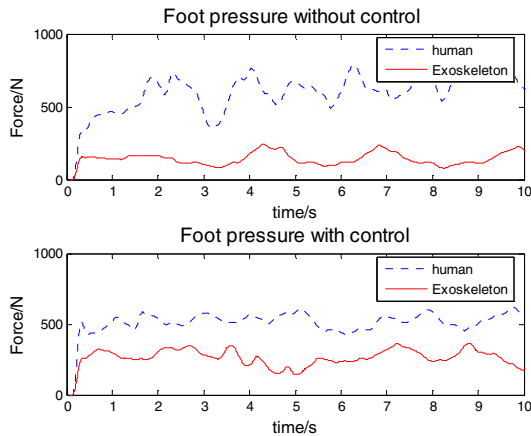
To validate the hardware and control method of ELEBOT discussed above, three experiments were conducted: load-carrying standing experiment, load-carrying walking experiment and up & down stairs experiment.



#### 4.1 Load-Carrying Standing Experiment

A load-carrying standing experiment was conducted with 30 kg payload. The foot pressures of operator and exoskeleton have been recorded respectively when exoskeleton was controlled and when exoskeleton was uncontrolled.

Fig. 6 shows a comparison between two groups of sensing records. When exoskeleton was uncontrolled, the foot pressure of operator is 200N higher than that when exoskeleton was controlled. This means about 200N load is supported by ELEBOT with above-mentioned hardware and control method.



**Fig. 6.** The comparison of the foot pressure of operator and exoskeleton during stance phase when exoskeleton was uncontrolled (above) and when it was controller



**Fig. 7.** The photos of ELEBOT in experiments. (a) ELEBOT supported 30 kg payload when the operator was standing. (b) ELEBOT went up stairs with 30 kg payload following the operator.

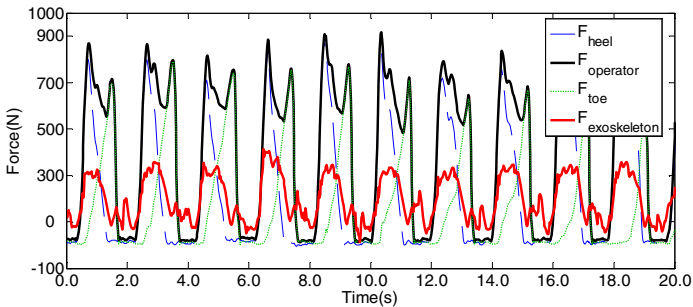
## 4.2 Up and Down Stairs Experiment

ELEBOT also could follow the operator's action to go up and down stairs (Fig. 7 (b) shows). This means the directly force feedback control is effective to the compliance of actuated knee joint.

## 4.3 Load-Carrying Walking Experiment

In this experiment, the operator walked with ELEBOT on a treadmill at a speed about 3.2 km/h, while the step length was about 0.45 m.

Fig. 8 shows the GRFs acquired from sensing shoe during this walking experiment. The blue long-dashed line is the forces on the heel of operator, the green short-dashed line is the forces on the toe of operator, the black solid line is the GRFs of operator and the red solid line is the GRFs of exoskeleton. This result indicated that about 300N payload was transferred to the ground via exoskeleton when ELEBOT was walking on level-ground with 30 kg payload under aforesaid control algorithm.



**Fig. 8.** The GRF reading from sensing shoe during the walking experiment with 30 kg payload. The operator was walking with ELEBOT at speed about 3.2 km/h, and 0.45 m step length.

## 5 Conclusion

While there is still significant work remaining, a leg exoskeletal prototype called ELEBOT has been developed. Entire system is mostly made up of pseudo-anthropomorphic mechanical structure, a kind of efficient, low-flow pump-valve hybrid controlled hydraulic actuator, a group of sensors and controller based on MCU. A knee torque estimate algorithm based on GRF is also discussed. Experiment indicted ELEBOT could walk while carrying its own weight and producing its own power. Walking experiment showed ELEBOT could support about 30 kg additional payload, walk at speeds up to 3.2 km/h, and follow the operator's maneuvers. The experiment results also indicated that application of force feedback PD controller to the knee joint control of exoskeleton is viable.

Further research will investigate following aspects: researching on flexible joint control algorithm and balance control method in heavy load condition.

**Acknowledgements.** This work was supported by the National Natural Science Foundation of China (Grant No.50775072).

## References

1. Herr, H.: Exoskeletons and orthoses: classification, design challenges and future directions. *Journal of Neuroengineering and Rehabilitation* 6, 21 (2009)
2. Kawamoto, H., Sankai, Y.: Power assist system HAL-3 for gait disorder person. In: *Computers Helping Person with Special Needs*, pp. 19–29. Springer, Heidelberg (2002)
3. Kazerooni, H., Steger, R.: The Berkeley lower extremity exoskeleton. *Transactions of the ASME, Journal of Dynamic Systems, Measurements, and Control* 128, 14–25 (2006)
4. Rose, J., Gamble, J.G.: *Human Walking*, 2nd edn. Williams & Wilkins, Baltimore (2004)
5. Harman, E., Han, K., Frykman, P., Pandorf, C.: *The Effects of Backpack Weight on the Biomechanics of Load Carriage*. Technical Report, Natick, Massachusetts, pp. T100–T117 (2000)
6. Mohammed, S., Amirat, Y.: Towards Intelligent Lower Limb Wearable Robots: Challenges and Perspectives – State of the Art. In: *IEEE International Conference on Robotics and Biomimetics*, pp. 312–317. IEEE Press, Bangkok (2009)
7. Zhen, J., Zhao, S., Wei, S.: Adaptively Fuzzy Iterative Learning Control for SRM Direct-drive Volume Control Servo Hydraulic Press. *Journal of Central South University of Technology* 17(2), 316–322 (2010)
8. Cao, H., Ling, Z., Zhu, J., Wang, Y., Wang, W.: Design frame of a leg exoskeleton for load-carrying augmentation. In: *IEEE International Conference on Robotics and Biomimetics*, Guilin, China, pp. 426–431 (2009)
9. Hurst, J.W., Rizzi, A.A.: Series Compliance for an Efficient Running Gait. *IEEE Robotics & Automation Magazine* 3, 42–51 (2008)
10. Kazerooni, H., Huang, L., Steger, R.: On the Control of the Berkeley Lower Extremity Exoskeleton (BLEEX). In: *IEEE International Conference on Robotics and Automation*, Barcelona, Spain, pp. 3469–3476 (2005)
11. Veneman, J.F., Ekkelenkamp, R., Kryudhof, R., Van der Helm, F.C.T., van der Kooij, H.: Design of a Series Elastic- and Bowdencable-based actuation system for use as torque-actuator in exoskeleton-type training. In: *9th IEEE International Conference on Rehabilitation Robotics*, Chicago, Illinois, pp. 496–499 (2005)

# Globally Optimal and Region Scalable CV Model on Automatic Target Segmentation

Huijie Fan<sup>1,2</sup>, Wei Dong<sup>1</sup>, and Yandong Tang<sup>1</sup>

<sup>1</sup> State Key Laboratory of Robotics, Shenyang Institute of Automation,  
Chinese Academy of Science, Shenyang 110016, China

<sup>2</sup> Graduate University of the Chinese Academy of Science, Beijing 100049, China  
{fanhuijie, ytang}@sia.cn

**Abstract.** This paper proposes a new Globally Optimal and Region Scalable Chan-Vese model (GRCV model), which combines the advantages of both local and global intensity information. Intensity information in local regions is emphasized by adding a kernel function in the data fitting term, which thereby enables the proposed model to extract the fine structure in local region. The global information is also considered by adding one data fitting term in Piecewise Constant (PC) model to guarantee that proposed model can successfully segment complete target regions from intensity inhomogeneous images. We give the selection rules of combining the different terms under different target and background intensities: when the target is intensity inhomogeneous while the background is homogeneous, we choose the data fitting term which effects on the outside region of the evolution contour; conversely, we choose the other term. Experiments have been done on different images to compare the effectiveness of our methods with that of the classical PC model, Li's Local Binary Fitting (LBF) model, and Wang's Local and Global Intensity Fitting (LGIF) model. Comparison results show that our model obtains more satisfactory segmentation results. Moreover, it is robust to the curve initialization and noise in images.

**Keywords:** Image Segmentation; Level Set Method; Active Contour; Intensity Inhomogeneous; Globally Optimal and Region Scalable CV Model.

## 1 Introduction

Image segmentation in image process can be formulated as an energy minimization problem and the active contour technique [1-5] provides a convenient framework to solve this problem iteratively. The existing active contour models can be classified as edge-based models [1, 2] and region-based models [3-5]. Edge-based models use a forcing function, computed by applying a differential operator over local edge information, to attract the evolving contour toward the object boundaries. Therefore, these kinds of model are sensitive to image noise and must be used in combination with image smoothing.

Chan and Vese proposed a Piecewise Constant (PC) model [4], which is a region-based active contour model without using the explicit edge detector. The PC model

assumes that the given image is consists of two regions, the objects to be detected and the background, and image intensities of two regions are statistically homogeneous (roughly a constant). By incorporating region-based information into their energy functional as an additional constraint, this model has large convergence range and flexible initialization. Compared with edge-based models, region-based active contour models have better performance in the presence of weak boundaries, stronger anti-noise performance, and weaker dependence on the initial curve. But most of them rely on intensity homogeneous inside each region.

Recently, Li et al. proposed a Local Binary Fitting (LBF) model [6, 7] by adding a kernel function in the data fitting term to overcome intensity inhomogeneous. This model draws upon accurate local intensity information, which enables it to deal with the segmentation problem in intensity inhomogeneous images. It can also detect the weak boundaries of much concave objects in images. However, this model focuses too much on local information, so it is sensitive to the position of initial curve, which causes the wrong segmentation results (shown in Fig. 3(d)) in some conditions. Wang et al. [8, 9] proposed an improved Local and Global Intensity Fitting (LGIF) model for brain MR image segmentation, which combines the data fitting term of the PC model with the LBF model. It collects the advantages of both local and global intensity information and works well on medical images. But when the boundaries of the targets are very weak, the segmentation results are incomplete (shown in Fig. 2(e) and Fig. 3(e)).

We found that the segmentation results are more reasonable when only one data fitting term is added in the LBF model under some conditions. We propose a Globally Optimal and Region scalable CV model (GRCV model) in which both local information and global information of the image are concerned by adding one data fitting term of the PC model into the energy functional of the LBF model. We give the rules to combine different terms under different intensity distribution: When the target is intensity inhomogeneous while the background is homogeneous, the data fitting term effecting on the outside region of the evolution contour should be chosen; conversely, the one effecting on inside region should be chosen. Some experiments on our method have been done on the synthetic and the real images and the experimental results are compared with those of the PC model, the LBF model and the LGIF model. The comparison shows that our method is more satisfactory than the others in extracting the boundaries.

The rest of this paper is organized as follows. Section 2 first presents the well-known region-based PC model [4] and the region-scalable LBF model [6] proposed by Chunming Li et al, followed our model combining one data fitting term in the PC model with the LBF model. The implementation and the comparison results are discussed in Section 3. The conclusion can be found in Section 4.

## 2 Depiction of the Models

### 2.1 Piecewise Constant Model

Tony Chan and Luminita Vese proposed a region-based PC model [4] to detect objects in an image, which is based on piecewise approximation of the well-known Mumford-Shah functional [10]. Instead of using edge detector for stopping the curve

evolution, the PC model uses region information (e.g. the intensity of image) and tries to minimize the energy functional of intensity variances within the target region and the background region. Therefore this model has stronger anti-noise performance and weaker dependence on the initial curve. The PC model has a level set formulation, which increases the dimensionality of the function to a higher one, but has an advantage of handling topological changes of the closed curves. Interior contours can be automatically detected using level set formulation.

Let  $\Omega$  be a bounded open subset of  $\mathbb{R}^2$  and  $C$  be an evolving curve denoting the boundary of the open subset  $\omega$  of  $\Omega$  (i.e.  $C=\partial\omega$ ). Let  $I: \Omega \rightarrow \mathbb{R}$  be a given image. In the level set method, the curve  $C$  is represented implicitly by the zero level set of a Lipschitz function  $\phi: \Omega \rightarrow \mathbb{R}$  (called level set function) [11]:

$$\begin{cases} C = \partial\omega = \{(x, y) \mid \phi(x, y) = 0\} \\ \text{inside}(C) = \omega = \{(x, y) \mid \phi(x, y) > 0\} \\ \text{outside}(C) = \Omega \setminus \omega = \{(x, y) \mid \phi(x, y) < 0\} \end{cases} \tag{1}$$

where  $(x, y)$  denotes the image coordinate. Assume that the image  $I$  consists of two regions which are approximately piecewise constant intensities and the object to be detected is one of the two regions. The final zero level set  $\{(x, y) \mid \phi(x, y) = 0\}$  corresponding to the flame front boundary. The ‘‘fitting energy’’ functional  $E$  of CV model is defined as follows:

$$\begin{aligned} E^{CV}(\phi, c_1, c_2) &= \lambda_1 \iint_{\Omega} \|I - c_1\|_{L^2}^2 H(\phi) \, dx dy \\ &+ \lambda_2 \iint_{\Omega} \|I - c_2\|_{L^2}^2 (1 - H(\phi)) \, dx dy + \nu \iint_{\Omega} \|\nabla H(\phi)\|_{L^2} \, dx dy \end{aligned} \tag{2}$$

where  $\lambda_1, \lambda_2$ , and  $\nu$  are non-negative weighting parameters, and  $H(\phi)$  is the Heaviside function of  $\phi$ .  $c_1$  and  $c_2$ , update with the evolution of the level set function  $\phi$ , are the mean intensities of the region  $I^+$  and  $I^-$  which are respectively the region inside and outside of the evolution contour. The first two items of  $E$  are called data fitting term, which measures the intensity variances of  $I^+$  and  $I^-$ . The last term is called the regularization term, which regularizes the curve length and smoothes the curve according to local curvature information. Segmentation is achieved by minimizing the energy functional in (2).

### 2.2 Globally Optimal and Region Scalable CV Model

In [6], Chunming Li et al. proposed a region based active contour model using intensity information in local regions. The key idea is to add a kernel function with a scale parameter into the data fitting term, which allows the use of intensity information in regions at a controllable scale. For a given point  $\mathbf{x}$  in image, the local intensity fitting energy is defined as follows:

$$e_{\mathbf{x}}^{Fit}(\phi, f_1, f_2) = \sum_{i=1}^2 \lambda_i \int_{\Omega} K_{\sigma}(\mathbf{x} - \mathbf{y}) \|I(\mathbf{y}) - f_i(\mathbf{x})\|_{L^2}^2 M_i(\phi(\mathbf{y})) \, dy \tag{3}$$

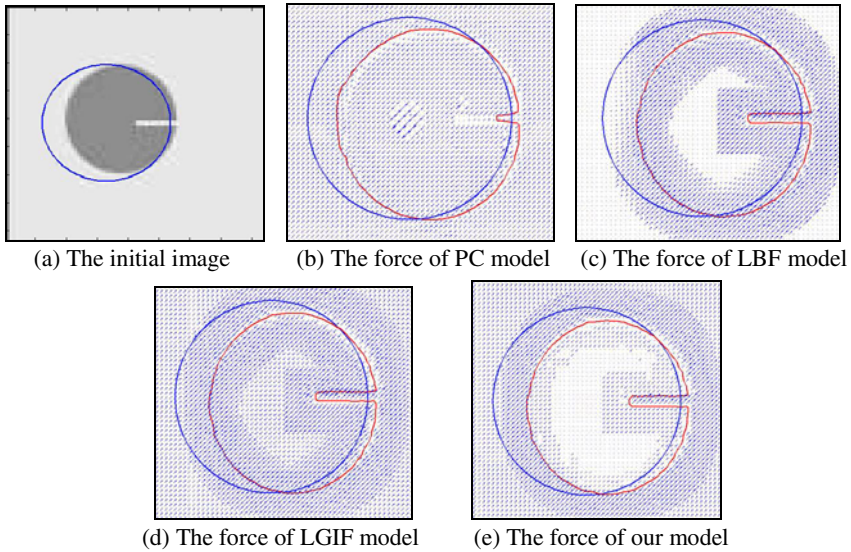
where  $f_1(\mathbf{x})$  and  $f_2(\mathbf{x})$  are two values that approximate image intensity in  $I^+$  and  $I^-$  respectively.  $M_1(\phi) = H(\phi)$  and  $M_2(\phi) = 1-H(\phi)$ . The intensities  $I(\mathbf{y})$  are a local region centered at the point  $\mathbf{x}$ , whose size is controlled by the kernel function  $K_\sigma$  with a scale parameter  $\sigma > 0$ :

$$K_\sigma(\mathbf{u}) = \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-|\mathbf{u}|^2 / 2\sigma^2} \tag{4}$$

The following energy functional is defined to obtain the entire object boundary:

$$E^{Region}(\phi, f_1, f_2) = \int e_x^{Fit}(\phi, f_1, f_2) d\mathbf{x} + \nu \int \|\nabla H(\phi(\mathbf{x}))\|_{L^2} d\mathbf{x} + \mu \int \frac{1}{2} (\|\nabla \phi(\mathbf{x})\|_{L^2} - 1)^2 d\mathbf{x} \tag{5}$$

where the second term is length regularization term which keeps the curve smooth, and the last term is called level set regularization term which characterizes the deviation of the level set function form a signed distance function.



**Fig. 1.** forces comparisons of LBF model and classical PC model

We show the forces comparisons of the classical PC model, the LBF model, the LGIF model, and the GRCV model on the curve evolution in Fig. 1. The blue contours in Fig. 1 represent the initial curve, the red contours are the current curve after ten iterations, and the blue arrows represent the force vectors. Fig. 1(b)-(d) respectively show the force of the PC model, the LBF model, and the LGIF model. We can see that the PC model can hold the whole movement of the evolution curve, even at the places far away from object boundaries, so it can handle the curve moves quickly to the targets boundaries but can not get the detailed edges; The force of the LBF

model is of great value at the place near object boundaries, so the model can attract the contour towards object boundaries and detect the precise object boundaries, but it is sensitive to initial curve because the force is weak at the place far away from the evolution curve. Moreover, the LBF model is easy to become stuck in local minima. The forces of the PC model and LBF model cooperate the force of the LGIF model, which exists in the global region, will lead to incomplete segmentation when the targets have very weak boundaries.

Our model has the advantages of PC model and LBF model, and is different from LGIF model [8] which combines both two data fitting term in the PC model. Only one global term in Formula (2) is added. The energy functional of the GRCV model is defined:

$$E(\phi, f_1, f_2) = E^{Region} + k_i \int \|I(\mathbf{x}) - c_i\|_{L^2}^2 M_i(\phi(\mathbf{x})) d\mathbf{x}, \quad i=1 \text{ or } 2 \quad (6)$$

where the nonnegative parameter  $k_i$  has the same meaning as  $\lambda_i$  in (2). Parameterizing the gradient descent direction by an artificial time  $t \geq 0$ , the Euler-Lagrange equation can be derived by minimizing the energy functional  $E$  with respect to  $\phi$

$$\left\{ \begin{aligned} \frac{\partial \phi}{\partial t} &= \delta_\varepsilon(\phi) [-\lambda_1 e_1 + \lambda_2 e_2 + \nu \nabla \cdot (\frac{\nabla \phi}{|\nabla \phi|}) + (-1)^{i-1} k_i (I - c_i)^2] \\ &\quad + \mu (\nabla^2 \phi - \nabla \cdot (\frac{\nabla \phi}{|\nabla \phi|})) \\ c_i &= \frac{\int_{\Omega} I(\mathbf{x}) M_i^\varepsilon(\phi(\mathbf{x})) d\mathbf{x}}{\int_{\Omega} M_i^\varepsilon(\phi(\mathbf{x})) d\mathbf{x}}, \quad i=1,2 \\ f_i(\mathbf{x}) &= \frac{K_\sigma(\mathbf{x}) * [M_i^\varepsilon(\phi(\mathbf{x})) I(\mathbf{x})]}{K_\sigma(\mathbf{x}) * M_i^\varepsilon(\phi(\mathbf{x}))}, \quad i=1,2 \\ \phi(\mathbf{x}, t=0) &= \phi_0(\mathbf{x}) \end{aligned} \right. \quad (7)$$

where  $M_1^\varepsilon(\phi) = H_\varepsilon(\phi)$  and  $M_2^\varepsilon(\phi) = 1 - H_\varepsilon(\phi)$ .  $H_\varepsilon(\phi)$  and  $\delta_\varepsilon(\phi)$  are especially the smooth approximation of Heaviside function and Dirac function given in (9).  $e_1$  and  $e_2$  are the functions:

$$e_i(\mathbf{x}) = \int K_\sigma(\mathbf{y} - \mathbf{x}) |I(\mathbf{x}) - f_i(\mathbf{y})|^2 d\mathbf{y}, \quad i=1,2 \quad (8)$$

$$H_\varepsilon(\phi) = \frac{1}{2} (1 + \frac{2}{\pi} \arctan(\frac{\phi}{\varepsilon})), \quad \delta_\varepsilon(\phi) = \frac{dH_\varepsilon(\phi)}{d\phi} = \frac{1}{\pi} \frac{\varepsilon}{\varepsilon^2 + \phi^2} \quad (9)$$

When the image contains the intensity inhomogeneous targets and the homogeneous background, the second data fitting term in PC model should be chosen in order that the energy functional corresponding to the region outside of the contour is minimal; conversely, the first data fitting term should be chosen. Fig. 1(e) shows the force of our model using the second data fitting term. The force outside of the contour will



pull the evolution curve close to the boundary and finally get the entire target region. The segmentation results are more satisfying than those of the LGIF model in the experiments.

### 3 The Experiment Results

The proposed method has been tested on synthetic and real images from different modalities. We compare our model with the PC model [4], Li’s LBF model [6] and Wang’s LGIF model [8]. The parameters are fixed in all experiments:  $\lambda_1 = \lambda_2 = k_1 = k_2 = 0.03$ ,  $\nu = 60$ ,  $\mu = 1$  and  $\sigma = 4$ .

We choose the high speed Planar Laser Induced Fluorescence (PLIF) images [12, 13] in the experiments. Statistical characteristics of flame surface shapes and flame front curvatures provide important information for understanding turbulence and chemistry interaction phenomena such as burning rate of the mixture, flame stabilization and the production of pollutant. The boundary detection of flame front, the main information source we can acquire, becomes an essential step. The intensities of the flame burning region is extremely inhomogeneous, and the long thin cracks appeared in flame surface have very weak boundaries (Fig. 2(a) and Fig. 3(a) show the gradient maps). So it is difficult to segment the crack regions precisely and extract the flame surface completely from the PLIF images.

The background is single and the flame surface to be segmented is intensity inhomogeneous, so we combine the second data fitting term of the PC model corresponding to the background region with the LBF model (i.e.  $i=2$  in (6)), it will generate an external force to pull the evolution contour close to the flame front boundary. Experiments have been done on both raw and de-noising PLIF images.

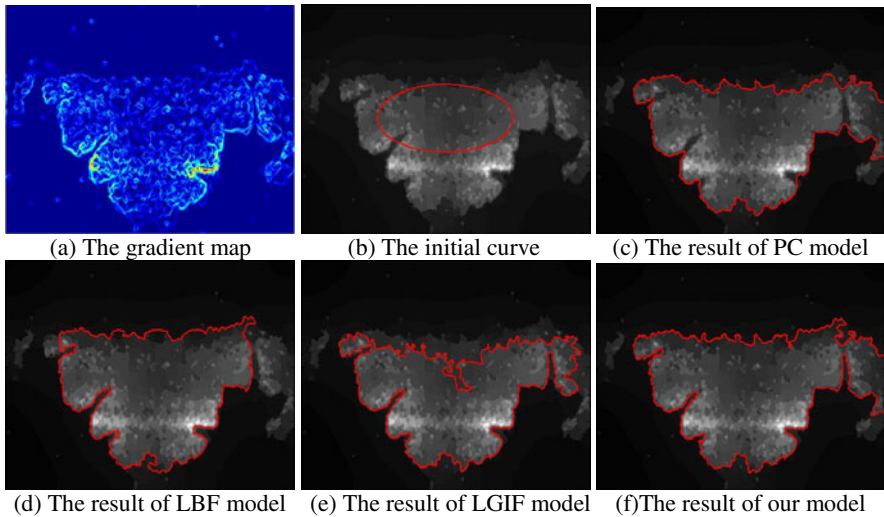
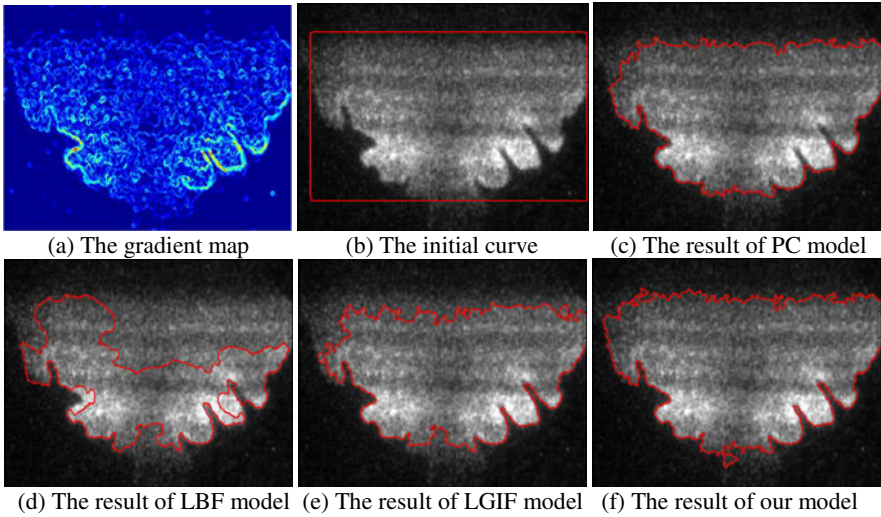


Fig. 2. The Flame front detection results on a de-noising PLIF image

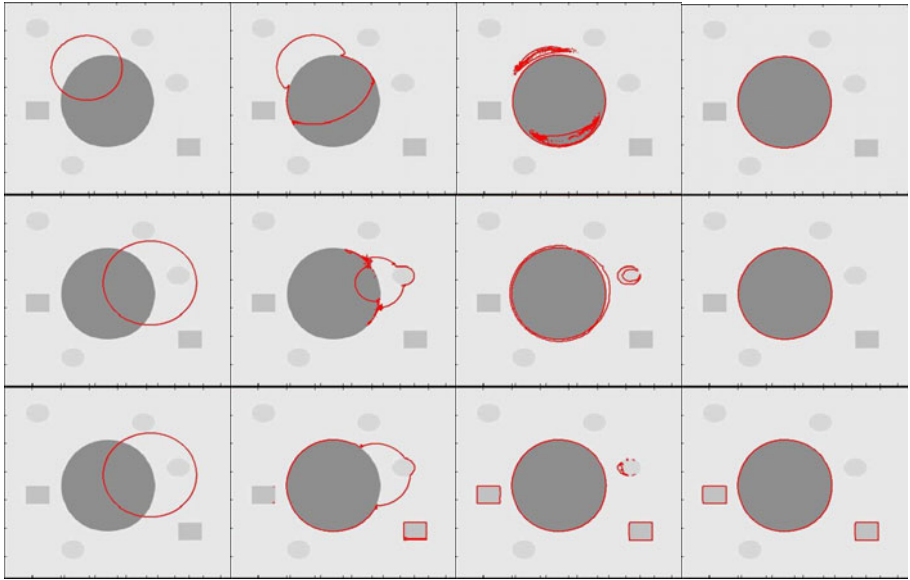
Fig. 2 gives the experiment results on a de-noising PLIF image disposed by the non-linear diffusion filtering [14, 15] whose gradients of flame front in the top and the bottom are very weak, as shown in Fig. 2(a) and Fig. 2(b) shows the de-noising image after non-linear diffusion filtering and the elliptical initial curve. We can see that part of the crack boundary is very weak. Fig. 2(c)~(f) are the results of the classical PC model, LBF model, and LGIF model. The PC model segments the whole flame surface but does not detect the crack boundary. The LBF model successfully extracts the crack boundary, but it misses parts of flame surface because it focuses too much on local information. The LGIF model cannot detect the entire flame surface either. Our algorithm can detect both the crack boundary and the complete flame surface.



**Fig. 3.** Flame front detection results on raw PLIF images

Fig. 3 shows the experiment results on a raw PLIF image. The intensity distribution of the flame surface is inhomogeneous: several brighter areas are scattered in image. To testify the insensitivity of our method to different initial curves, a red rectangle initial curve is used in Fig. 3(b), which is different from that in Fig. 2(b). The results of the methods are shown in Fig. 3(c)~(f): the PC model detects the entire flame surface but fails to extract the edge of crack region; the LBF model has the problem of over-segmentation, and it is easy to be stuck in local minima. Our model achieves the better result than that of LGIF model because the external force, generated by the global term acting on outside contour, pulls the contour close to the real boundary. The experiment also demonstrates that our algorithm is robust to image noise.

Another experiment is done on a synthetic image, in which the big circle is target and the small objects around the circle belong to the inhomogeneous background, as is shown in Fig. 4. We combine the first data fitting term in the PC model with the LBF model (i.e.  $i=1$  in (6) and (7)). because the first data fitting term is corresponding to the target region, and it will generate an internal force to push the evolution contour



**Fig. 4.** Object detection with different initial curves. The first two rows are the curve evolution processes of our model and the last row show that of LGIF model; the curve evolution process is shown from the initial contour (in the first column) to the final contour (in the fourth column).

to the target boundary. The first two rows are the curve evolution processes of our model. Given the two different initial states, the target is successfully segmented by our method. The last row is the curve evolution process of the LGIF model. The LGIF model fails in getting rid of two surrounding objects in the background.

## 4 Conclusion

In this paper, we propose a Globally Optimal and Region Scalable Chan-Vese model (GRCV model), which considers both local and global intensity information. The proposed model draws up intensity information in local regions at a controllable scale; therefore it is able to extract the fine structure in local region, such as cracks which are long thin areas with quite weak boundaries in PLIF images. The global information is also considered in energy functional of GRCV model, so our algorithm can successfully segment the complete target region from intensity inhomogeneous images. We compare our model with the classical PC model, Li's LBF model, and Wang's LGIF model on different intensity inhomogeneous images. The experimental results show that our model which takes into accounts both global and local features of images can obtain more satisfactory segmentation results. Moreover, it is robust to the curve initialization and noise in images.

**Acknowledgments.** This work has been funded by Natural Science Foundation of China (Grant No. 50876110, 60871078 and 60835004). We would like to thank Zhu Linlin, Sun Jing, Fan Baojie and Du Yingkui for their advice and support.

## References

1. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. *International Journal of Computer Vision* 1, 321–331 (1987)
2. Xu, C., Prince, J.: Snakes, shapes, and gradient vector flow. *IEEE Transaction on Image Processing* 7(3), 359–369 (1998)
3. Ronfard, R.: Region-based strategies for active contour models. *International Journal of Computer Vision* 13, 229–251 (1994)
4. Chan, T., Vese, L.: Active contours without edges. *IEEE Transaction on Image Processing* 10, 266–277 (2001)
5. Vese, L., Chan, T.: A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision* 50, 271–293 (2002)
6. Chunming, L., Chiuyen, K., Gore, J.C., Zhaohua, D.: Implicit active contours driven by local binary fitting energy. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2007)
7. Chunming, L., Chiuyen, K., Gore, J.C., Zhaohua, D.: Minimization of region-scalable fitting energy for image segmentation. *IEEE Transactions on Image Processing* 17, 1940–1949 (2008)
8. Li, W., Chunming, L., Quansen, S., Deshen, X., Chiuyen, K.: Brain MR Image Segmentation Using Local and Global Intensity Fitting Active Contours/Surfaces. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I. LNCS*, vol. 5241, pp. 384–392. Springer, Heidelberg (2008)
9. Li, W., Chunming, L., Quansen, S., Deshen, X., Chiuyen, K.: Active contours driven by local and global intensity fitting energy with application to brain MR image segmentation. *Computerized Medical Imaging and Graphics* 33(7), 520–531 (2009)
10. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.* 42, 577–685 (1989)
11. Osher, S., Sethian, J.: Fronts propagating with curvature dependent speed: algorithms based on the Hamilton-Jacobi formulation. *Journal of Computational Physics* 79, 12–49 (1988)
12. Kaminski, C.F., Hult, J., Aldén, M.: High repetition rate planar laser induced fluorescence of OH in a turbulent nonpremixed flame. *Applied Physics B* 68, 757–760 (2000)
13. Tanahashi, M., Murakami, S., Choi, G.M., Fukuchi, Y., Miyachi, T.: Simultaneous CH–OH PLIF and stereoscopic PIV measurements of turbulent premixed flames. *Proc. Combust. Inst.* 30, 1665–1672 (2005)
14. Malm, H., Hult, J., Sparr, G., Kaminski, C.F.: Non-linear diffusion filtering of images obtained by planar laser induced fluorescence spectroscopy. *Journal of the Optical Society of America A (JOSA A)* 17(12), 2148–2156 (2000)
15. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(7), 629–639 (1990)

# The Registration of UAV Down-Looking Aerial Images to Satellite Images with Image Entropy and Edges

Baojie Fan<sup>1,2</sup>, Yingkui Du<sup>1,2</sup>, Linlin Zhu<sup>1,2</sup>, and Yandong Tang<sup>1</sup>

<sup>1</sup> State Key laboratory of Robotics, Shenyang Institute of Automation,  
Chinese Academy of Science,  
110016 Shenyang, China

<sup>2</sup> Graduate School of the Chinese Academy of Science,  
100039 Beijing, China  
{fanbaojie, dyk, zhulinlin, ytang}@sia.cn

**Abstract.** In this paper, we propose a novel and efficient image registration algorithm between high resolution satellite images and UAV down-looking aerial images. The algorithm is achieved by a composite deformable template matching. To overcome the limitations of environment changes and different sensors, and to remain image information, we fuse the image edge and entropy features as image representation. According to the altitude information of the UAV, we can get the scales of the down-looking aerial images relative to the satellite images. In the following, we perform an effective search strategy in the satellite images to find the best matching position. Different experimental results show that the proposed algorithm is effective and robust.

**Keywords:** image registration, deformable template match; entropy image; image edge.

## 1 Introduction

Image registration has found applications in numerous real life applications such as remote sensing, medical image analysis, computer vision and pattern recognition [1]. Given two, or more images to be registered, image registration estimates the parameters of the geometric transformation model that maps a given image to the reference one. Geo-registration is a very useful technique, which can be widely used in UAV (Unmanned Aerial Vehicle) to navigate, or to geo-locating a target, or even to refine a map [2].

Feature-based registration methods have made great progress in dealing with aerial images [3,4,5] and aerial image sequences [6,2] in recent years. Tuo et al. [3] perform registration after modifying the images to fit a specified brightness histogram. The features are then detected and aligned. Yasein and Agathoklis [4] solve only for a similarity transformation, but use an iterative optimization where the points are weighted according to the current residual. Xiong and Quek [5] perform registration up to similarity transformations without explicitly finding correspondences. After detecting features in both images, an orientation is computed for each feature and all possible correspondences are mapped into a histogram according to the orientation

differences. The peak in the histogram is chosen as the rotation between the images. Scale is determined through the use of angle histograms computed with multiple image patch sizes and selecting the highest peak from the histogram. Niranjana et al. [6] build upon the work of Xiong and Quek in order to register images in an image sequence up to a homography. Lin et al. [2] concentrate on registering consecutive aerial images from an image sequence. They use a reference image (such as a map image) in order to eliminate errors that accumulate from local methods. Their two-step process first performs registration between images and then uses this as an initial estimate for the registration with the reference image.

Various works have been presented for the UAV pose estimation using a camera. A localization method matching digital elevation maps with aerial images has been suggested [7], and relative and absolute UAV localization methods by matching between satellite images and down-looking aerial images have been studied [8]. Caballero estimates a relative position of the UAV by calculating homographies among down-looking aerial scenes with the assumption that the ground is a plane [9][10]. Kil-Ho Son develops an UAV global pose estimation algorithm by matching the forward-looking aerial images with the satellite images [11].

The paper is organized as follows. In Section 2, we introduce the background of the proposed image registration algorithm. Then we describe the novel deformable template matching method in Section 3. The experimental results are presented in Section 4, which is followed by some conclusions in Section 5.

## 2 Background

Access to high resolution images for many areas of the world does not represent a problem any longer. So, it is feasible that we regard the high resolution satellite images provided by the up to date Google Earth software as the geo-referenced image. Besides, the UAV is equipped with down-looking vision systems and the laser altimeter. The viewpoint of the vision system is similar to that of the satellite, so we assume that the motion between satellite images and down-looking aerial images agrees to affine transformation model in Equation 1. This is convenient to the following matching.

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = A \begin{bmatrix} x_u \\ y_u \end{bmatrix} + T = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_u \\ y_u \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \quad (1)$$

Where  $(x_s, y_s)$ ,  $(x_u, y_u)$  are the coordinates in satellite image and down-looking aerial image respectively,  $A$  is the deformation matrix,  $T$  is the translation vector.

The laser altimeter can give the UAV altitude relative to the ground. The altitude can give a cue to calculate the scales of the down-looking aerial images relative to the satellite images. This is very important information for us. Scaling is performed converting the aerial image to the resolution of the reference image. The scale factor  $s$  is calculated using Equation 2 and it is different in  $x$  and  $y$  direction of the image plane since the aerial images used do not have squared pixels.

$$\begin{bmatrix} s_x \\ s_y \end{bmatrix} = \begin{bmatrix} \frac{1}{f_x} \\ \frac{1}{f_y} \end{bmatrix} H \times I_{res} \quad (2)$$

$H$  is the altitude of the UAV relative to the ground given by the laser altimeter.  $I_{res}$  is the resolution of the reference image provided by the Google Earth.

In the following, we perform a novel deformable image matching between geo-referenced image and down-looking aerial image to find the UAV position, and give the values of affine parameters.

### 3 Novel Deformable Template Matching

The geo-referenced and the video camera image are generally taken at different time. It can be months or years, the illumination conditions will differ. Therefore, it is necessary to choose the features which are robust to the illumination changes. A Sobel edge detector is applied to both the geo-referenced image and the image taken from the on-board video camera. The choice of using edge features derives from the fact that the edges are quite robust to environmental illumination changes.

Another important factor to be considered is the altitude of the UAV from the ground. The higher the UAV flies, the more structure from the environment can be captured. It means that image registration is more reliable at higher altitude. Considering that small details change quite fast (e.g. car moving on the road) while large structures tend to be more static (e.g. roads, buildings...), flying at higher altitude makes the registration more robust to small dynamic changes in the environment. Besides, the geo-referenced and the aerial image are captured with different sensors. We must find the features which are insensitive to many issues in multi-sensor matching while retaining much image information. The entropy image is the best selection, which is proved by Clark F. Olson in [12]. The entropy image is also invariant to the illumination changes, which is a strong support to the edge feature. But the obtained entropy image is influenced by the window size. The edge feature can overcome this drawback.

In this paper, we fuse the edge and entropy of the template as its representatives, the both features are complementary, which are robust to environment changes, different sensors and retain image information. We aim to maximize the normalized correlation between high resolution satellite images and down-looking aerial images to find the best matching position. In our proposed algorithm, there are two terms in the deformable template matching function, as shown in Equation 3. The flowchart is shown in Fig. 1.

$$NCC_{total} = (NCC_{edge} + NCC_{entropy}) / 2 \quad (3)$$

In general, the deformable template matching involves rotating and rescaling the aerial image according to the pose parameters. The scale parameter can be estimated in Section 2. We can obtain the rotation parameter with the best matching position by searching in the geo-referenced image.

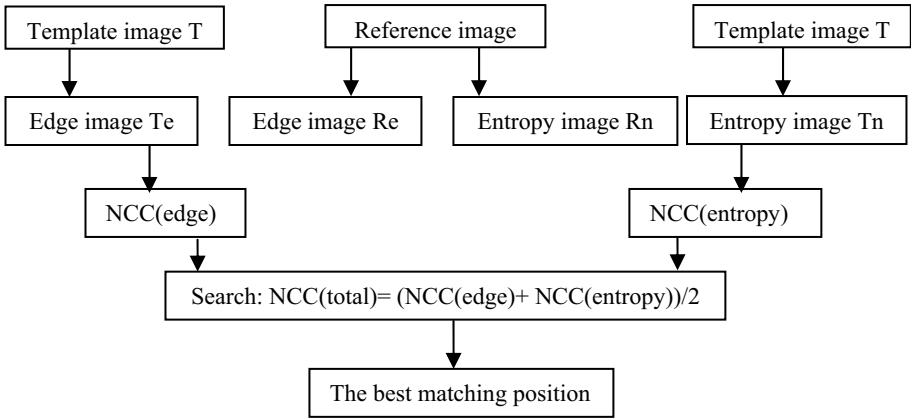


Fig. 1. The flow chart of our proposed deformable template matching algorithm

### 3.1 The Edge Operation

After the color aerial images are converted to the gray images, a median filter is applied to remove small details which are visible in the aerial image but not visible in the reference one. It is also capable of preserving the edges sharp when removes the details. After filtering, the Sobel edge detector is applied. Then scaling is performed converting the aerial image to the resolution of the reference image. The calculation of the scale parameters refers to the method in Section 2. According to the reference image, it is converted into gray images, and the Sobel edge detector is applied. The aerial and reference edge images are saved for the following deformable template matching.

### 3.2 The Entropy Image

For a discrete random variable  $A$ , with the marginal probability distribution  $p(A)$ , the entropy is defined as

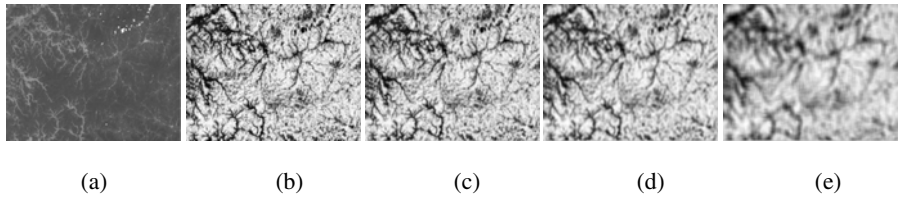
$$H(A) = -\sum_A p(A) \log p(A) \tag{4}$$

Note that  $0 \times \log 0$  is taken to be zero, since

$$\lim_{x \rightarrow 0} x \log x = 0 \tag{5}$$

We apply an entropy transformation to both the aerial image and reference image as follows. For each image location  $(x, y)$ , we examine the intensity values in an image window centered  $(x, y)$  with the size  $k \times k$ . The intensities are histogrammed and the entropy for the window is computed according to the Equation (4). In practice, it is useful to smooth the histogram prior to using the Equation (4). For efficiency, we use a histogram with 128 bins and smooth the histogram using a Gaussian window.





**Fig. 2.** Entropy images at different scales. (a) The original image (b)By 5\*5window (c)By 7\*7window (d)By 11\*11window (b)By 21\*21 window.

Fig. 2 shows an example of the entropy images generated at different scales. We can see that the entropy transformation retains the amount of local variation information in image. As the size of the window increases, the local entropy is spread and smoothed over a larger area, which provides an important basis for the following deformable template matching.

### 3.3 The Efficient Search

In order to get the optimism matching between the aerial image and the reference image, we use the FFT to quickly determine the normalization values for each translation of the aerial image. The search strategy is as follows.

According to the deformable template matching, we only know the scale of aerial images in the reference image, the rotation is unknown. We can search in the reference image to find the suitable angle with the best matching. The range of the rotation belongs to  $[0, 2\pi)$ , we divide the range equally into 36 parts, that is,  $\theta_i = 0^\circ, 10^\circ, 20^\circ \dots 35^\circ$ . With the obtained scale, each angle  $\theta_i$  corresponds to a template  $T_i$ , in the following, we find the best matching position of the template  $T_i$  in the reference image, and give the normalized correlation value  $NCC_i$ . We calculate the average of the adjacent normalized correlation values:  $NCC_{i-1}$ ,  $NCC_i$ , and  $NCC_{i+1}$ . The maximum of the average values is the optimism matching with the interval  $[\theta_{i-1}, \theta_{i+1}]$ . The interval  $[\theta_{i-1}, \theta_{i+1}]$  is subdivided into smaller sections recursively to get the interval closer to the optimism matching position. The operation is repeated until the interval converges to a small neighborhood.

## 4 Experiments

In this section, the performance of the proposed algorithm is evaluated for the different reference images and the down-looking aerial images. The high resolution satellite images from the Google earth are as the reference images in this paper. The UAV aerial images with the altitude information are provided. In order to calculate the scales, we calibrate the used camera with the calibration toolbox of Matlab beforehand. In the first experiment, the aerial images are obtained by the simulation motion platform of the UAV vision system (shown in Fig. 3(a)). In other experiments, the aerial images are captured by video camera fixed on the UAV (shown in Fig. 3(b)).

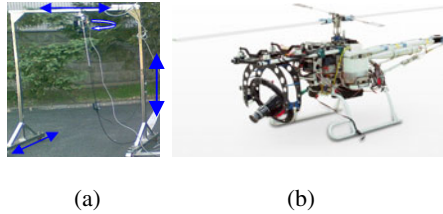
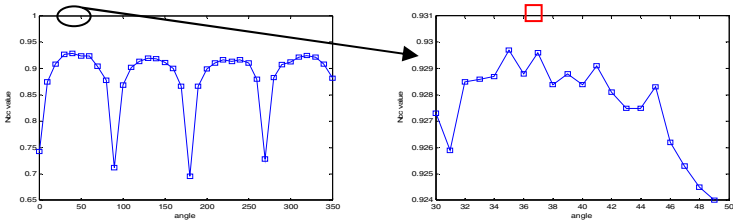


Fig. 3. The UAV and its vision simulation motion platform

In the first experiment, the reference image is obtained by the Canon digital camera, which is about forest sand table from 1.7 meters. An image of a tree in forest sand table is as aerial image in this experiment from 0.24 meters. The two images are from different sensors and altitudes. The search process and the matching result are shown in Fig. 4. It is encouraging.



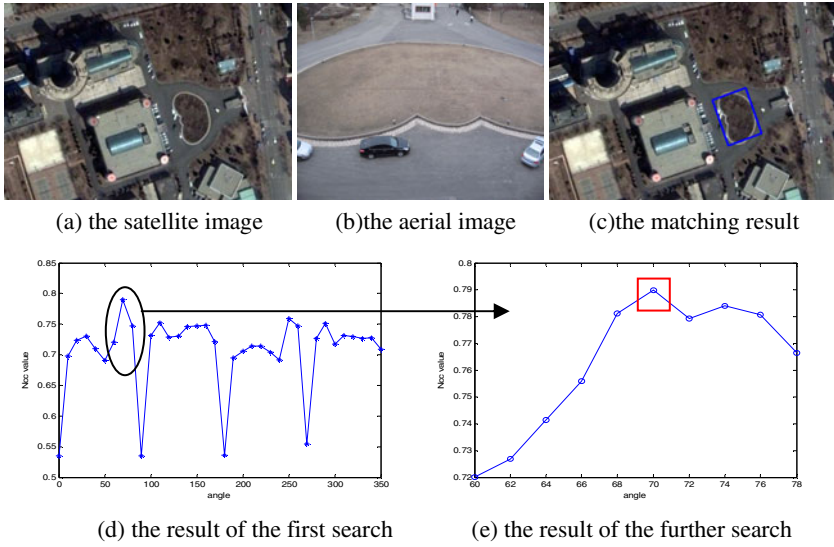
(a) the tree image (b) the forest sand table image (c) the matching result



(d) the result of the first search (e) the result of the further search

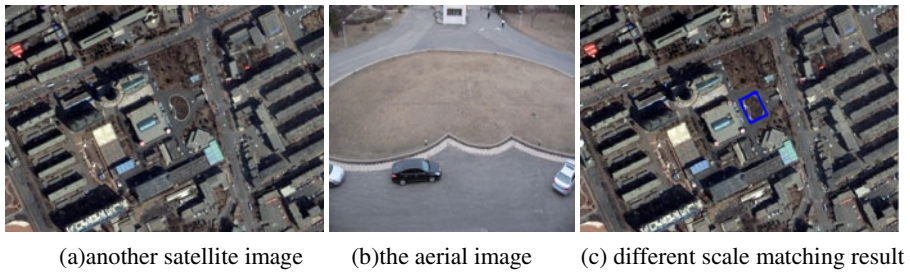
Fig. 4. A simulation experiment about forest sand table

A satellite image of Shenyang institute from the Google Earth is acquired from 300 meters above ground on Apr.2nd, 2009. The aerial image is obtained at 15 meters on Mar. 30th, 2010.  $s_x, s_y$  are gotten by Equation (2) with the help of the obtained camera internal parameters above. The two images are taken by different sensors and different environments. Firstly, we do the brute force search with the search strategy in Section 3.3. The normalized correlation values with different angles are shown in Fig. 5(d). The largest average of the normalized correlation values with the corresponding interval is at [60, 80]. The search continues until finding the suitable angle with the best matching, this is shown in Fig. 5(e). We can see from Fig. 5(c) that the experimental result of our proposed algorithm is satisfactory.

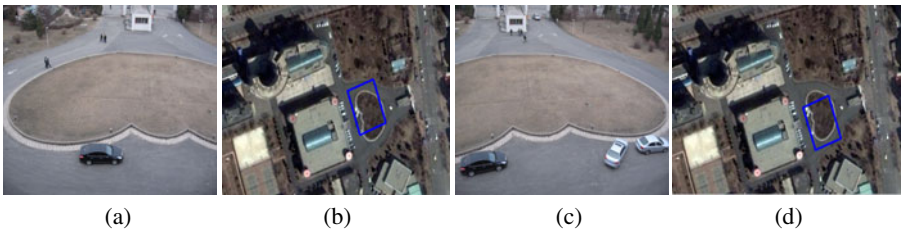


**Fig. 5.** The matching result from the images of Shenyang institute

With the same aerial images and different satellite image, the satellite image is acquired from the higher altitude, so the scale  $s_x, s_y$  is calculated again. We can get another encouraging result of deformable template matching as shown in Fig.6



**Fig. 6.** The matching result from different scale satellite images of Shenyang institute



**Fig. 7.** The matching results from different aerial images and same satellite image of Shenyang institute (a) the aerial image1, (b) the corresponding matching result from the aerial image1, (c) the aerial image2, (d) the corresponding matching result from the aerial image2

With the same satellite image in Fig.5 and different aerial image, another experiment is done to verify the robustness of the novel deformable template matching algorithm shown in Fig. 7.

From the above experimental results, we can see that our proposed algorithm is efficient and robust. Because, the viewpoint of the UAV camera is not completely down-looking and forward-looking, there are small errors in the matching. we will solve the problem in the future works.

## 5 Conclusions

In this paper, we develop a novel image registration algorithm between high resolution satellite images and UAV down-looking aerial images. The algorithm is achieved by a composite deformable template matching. The edge and the entropy of images are used to solve the problem brought by the changes of the illumination and different sensors. The combination of the edge the entropy also can retain most of the image information. The results of experiments on different types of images show that the proposed algorithm is effective and robust.

In the future, we will improve the efficient of our algorithm by updating the search strategy in composite deformable template matching, and extend the applications of the method for UAV pose estimation and UAV image mosaicing.

**Acknowledgments.** This work was supported by Natural Science Foundation of China (No.60805046 and 60835004).

## References

1. Lisa, G., Brown: A Survey of Image Registration Techniques. *ACM Computing Surveys (CSUR)* 24(4), 325–376 (1992)
2. Lin, Y., Yu, Q., Medioni, G.: Map-enhanced UAV Image Sequence Registration and Synchronization of Multiple Image Sequences. In: *IEEE Conference on Pattern Recognition*, pp. 1–7 (2007)
3. Tuo, H., Zhang, L., Liu, Y.: Multisensor Aerial Image Registration Using Direct Histogram Specification. In: *Proceedings of the IEEE International Conference on Networking, Sensing and Control*, pp. 807–812 (2004)
4. Yasein, M.S., Agathoklis, P.: A Robust, Feature-based Algorithm for Aerial Image Registration. In: *Proceedings of the IEEE International Symposium on Industrial Electronics*, vol. 4, pp. 1731–1736 (2007)
5. Xiong, Y., Quek, F.: Automatic Aerial Image Registration without Correspondence. In: *Proceedings of the 4th International Conference on Computer Vision Systems* (2006)
6. Niranjan, S., Gupta, G., Mukerjee, A., Gupta, S.: Efficient Registration of Aerial Image Sequences without Camera Priors. In: Yagi, Y., Kang, S.B., Kweon, I.S., Zha, H. (eds.) *ACCV 2007, Part II. LNCS*, vol. 4844, pp. 394–403. Springer, Heidelberg (2007)
7. Kumar, R., Sawhney, H.S., Asmuth, J.C., Pope, A., Hsu, S.: Registration of Video to Geo-Referenced Imagery. In: *Fourteenth International Conference on Pattern Recognition*, vol. 2, pp. 1393–1400 (1998)

8. Sim, D.G., Park, R.H., Kim, R.C., Lee, S.U., Kim, I.C.: Integrated Position Estimation Using Aerial Image Sequences. *IEEE Trans. Pattern Anal. Machine Intell.* 24(1), 1–18 (2002)
9. Caballero, F., Merino, L., Ferruz, J., Ollero, A.: Improving Vision-based Planar Motion Estimation for Unmanned Aerial Vehicles through Online Mosaicing. In: *International Conference on Robotics and Automation*, pp. 2860–2865 (2006)
10. Caballero, F., Merino, L., Ferruz, J., Ollero, A.: Homography Based Kalman Filter for Mosaic Building. Application to UAV Pose Estimation. In: *International Conference on Robotics and Automation*, pp. 2004–2009 (2007)
11. son, K.-h., Hwang, Y., Kweon, I.-S.: UAV Global Pose Estimation by Matching Forward-looking aerial Images with Satellite Images. In: *The 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3880–3887 (2009)
12. Olson, C.F.: Image Registration by Aligning Entropies. In: *IEEE Conference on Pattern Recognition*, vol. 2, pp. 331–336 (2001)

# Fast 3D Time of Flight Data Segmentation Using the U-V-Histogram Approach

Thomas Schamm, Arne Rönnau, and J. Marius Zöllner

Intelligent Systems and Production Engineering,  
FZI Forschungszentrum Informatik, 76131 Karlsruhe, Germany  
{schamm,roennau,zoellner}@fzi.de

**Abstract.** For modern assistance systems, time of flight (tof) cameras play a key role in the perception of the environmental situation. Due to benefits of this sensor type, especially in combination with video sensors, it outperforms typical sensor systems, like stereo cameras. However, tof depth image segmentation is yet not solved satisfactorily. Common approaches use homogeneous constraints and therefore assume objects to be parallel to the camera plane. This consequently leads to segmentation errors. This paper proposes a fast segmentation algorithm for detecting distinct planes using statistical knowledge. By projecting the depth image data along the image axis, u-v-histogram images can be constructed. It is shown, that in the histogram images, 3D planes correspond to line segments. A fast approach for line segment extraction in the histogram images is used to find the relevant planes. The algorithm was successfully tested with real data under varying conditions and can be applied for different tof camera sensors.

## 1 Introduction

With the development of high-end robotic systems, directly cooperating in human centered environments, much effort was taken to make robotic systems more adaptable to dynamic environments. To operate successfully in dynamic environments, a key part is a robust perception of the environmental situation. As long as environment complexity is low, 2D laser scanners may be used. Radar systems, which provide 3D range measurements, are considered robust against environment influences, but their limited beam angle and resolution lead to difficulties in complex environments. Hence, stereo sensors, 3D laser scanners as well as time-of-flight (tof) cameras have proofed as reasonable sensors for complex environment perception.

Fast rotating 3D laser scanners like the Velodyne HDL-64E provide full 3D environment perception, but they are huge in size, have a low refresh rate and still are not yet fail-safe. Stereo camera systems are fast, compact, cheap and well investigated. Modern approaches in stereo image analysis provide fast computation of disparity images. Information provided by these systems can be either sparse or dense disparity images, depending on the used algorithms. However, measurement errors occur at homogeneous structures, where point correspondence analysis fails.

Thus, tof cameras propose to play a key role for systems, where sensors must be small, fast and robust. Their ability to acquire full 3D environment data with only a single light burst outperforms the more expensive laser scanners.

Up to now, research in robotic vision is in search of fast and robust methods, to acquire and organize environmental data. On the one hand, the sheer amount of data from the 3D point clouds is a challenge for real time algorithmic processing. On the other hand, data segmentation must be accomplished problem specific. For example, object manipulation tasks require to percept all details of an object, thus exact segmentation is of high relevance. Instead, for collision avoidance tasks, the primary goal is to identify distinct objects as well as drivable planes, which requires to extract relevant environmental structures.

In this work, we present a fast and robust approach for plane segmentation in 3D point clouds acquired from tof cameras on a mobile platform. The algorithm is based upon the u-v-histogram segmentation approach basically known from stereo disparity images. To be applied successfully, the histogram calculation must be adapted, because stereo disparity data behaves different compared to tof depth data. To extract relevant planes from the histogram images, a fast line pattern search is presented. Planes orthogonal to the ground are considered as relevant obstacles, while planes parallel to the ground represent possible floor planes.

At first, the depth image acquired by the camera must be enhanced to reduce noise. Second, the corresponding u-v-histogram images of the depth image are calculated. Next, the histogram images are searched for distinct planes. This search is realized with a modified Hough transformation and line pattern search algorithm. At last, based on the resulting lines segments, the full object dimension is estimated.

This paper is organized as follows: Section 2 describes previous work in 3D point cloud analysis. The used coordinate systems in this work are clarified in Section 3. Our proposed method for plane segmentation is then presented in Section 4. Section 5 shows results of our plane segmentation algorithm on experimental data for two different tof cameras. This work's conclusion and future research is presented in Section 6.

## 2 Previous Work

Typically, mobile robots work in undefined, unstructured environments. For these environments, point cloud segmentation algorithms are limited to model free approaches, mainly because assumptions about e.g. cylindric, cuboid or comparable object types hardly fit in the point cloud data.

For object detection and free space classification, the class of model free algorithms can be divided into ground plane based methods, mostly applied with occupancy grids, and range image based methods. While occupancy grids provide the opportunity of cell based data fusion of multiple sensors (e.g. [12]), 3D information is lost due to ground plane projection. These approaches are not suitable in non flat areas as well.

Range image based methods are applied, where data can be projected onto a virtual camera plane. An overview of segmentation algorithms for range images was already presented in the 90's by [3]. However, most algorithms are only applied in artificial environments. For unstructured environments, graph-based methods seem to be the most promising segmentation methods. In [4], a fast segmentation method is presented for segmenting large 3D point clouds. The data is achieved with rotating 3D laser scanners, hence noise in the distance information is low.

For tof cameras, however, due to sensor limitations, the provided 3D point clouds are noisy and measurement errors occur at object edges or during fast motion [5]. Currently, available algorithms for processing tof point clouds are rare and still not mature. Many algorithms filter noisy image data using local homogeneity criterion, as presented in [6,7]. Other algorithms make usage of 2D image segmentation algorithms to segment depth points, e.g. unseeded region growing [8]. However, these algorithms are limited in only segmenting planes parallel to the image plane, which consequently leads to segmentation errors of the floor plane or large objects not parallel to the image plane.

More advanced segmentation algorithms exist in the field of stereo disparity images. The construction of disparity histograms from stereo image pairs was first introduced by [9] for road plane extraction. [10] uses the proposed method for robust object detection as well. In this work we extend these approaches, such that they can be applied on tof image data.

### 3 Coordinate Systems

Fig. 1 depicts both relevant coordinate systems, the tof image coordinate system (ICS) and the world coordinate system (WCS). In the ICS, the position of a point in the image plane is given by its coordinates  $(u, v, d)$ , whereas  $d$  is the radial distance of the point to the optical center. In the WCS, the  $x$ -axis points forward, the  $y$ -axis to the left and the  $z$ -axis upwards. The intrinsic parameters are given by the tuple  $(f, p_x, p_y)$ , where  $f$  is the focal length of the lens and  $p_x, p_y$  are the principal points of the image plane. The extrinsic parameters are defined by  $(t_x, t_y, t_z, \phi, \theta, \psi)$ , giving the translation and rotation of the ICS into the WCS.

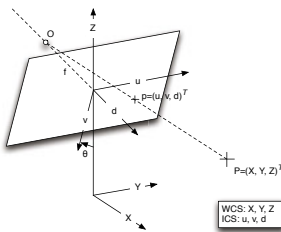


Fig. 1. Definition of coordinate systems

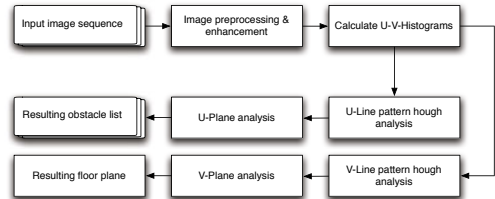


Fig. 2. Flowchart of algorithm

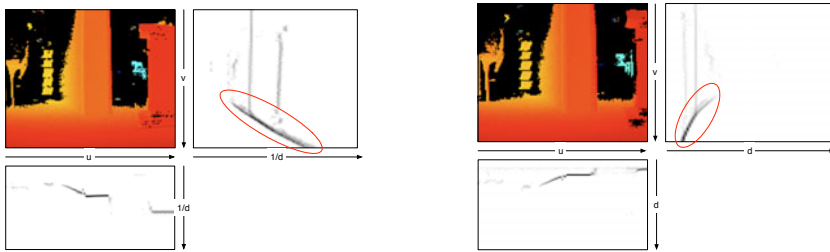


Using the pin-hole camera model, a projection of a point  $(X, Y, Z)$  on to the tof image plane is expressed by:

$$\begin{aligned} u &= f \frac{Y}{X} + p_x \\ v &= f \frac{Z}{X} + p_y \\ d &= \frac{X}{f} \sqrt{u^2 + v^2 + f^2} \end{aligned} \tag{1}$$

### 4 Proposed Method

The algorithm we propose can be separated into four steps. At first we have to preprocess the depth and intensity image pairs in terms of noise reduction and distance error correction. Next, the histograms along the horizontal (u) and vertical (v) image axis are calculated. Next, we use a modified line pattern Hough analysis to find distinctive lines in each histogram image, whereas each line represents a plane in the point cloud. At last, we determine vertical and horizontal dimensions of all planes by analyzing the depth image. Fig. 2 shows the flowchart of the proposed algorithm.



**Fig. 3.** False color tof image with corresponding u-/v-histograms, floor plane marked red. (left) Reciprocal construction. (right) Linear construction.

#### 4.1 Image Enhancement

The proposed method is robust against measurement noise by filtering image data with a low signal to noise level, and by using statistical methods. A common tof dataset consists of at least one image containing the distance information (called depth image) and one corresponding to the amount of reflected light (intensity image). While most tof cameras preprocess the captured images on chip, the intensity image still contains valuable information to enhance the provided depth data. As the signal to noise ratio of the depth data correlates with the intensity of modulated light reflected by the scene, distance data with low signal to noise ratio is filtered by analyzing the intensity value:

$$d = \begin{cases} - , & \text{if intensity} < T_i \\ d , & \text{otherwise} \end{cases} \tag{2}$$

where  $T_i$  is a sensor-specific threshold, which can be obtained by sensor calibration [8]. The subsequent segmentation steps use histograms to estimate the probability density of the depth image, such that measurement noise not filtered due to its signal to noise ratio is suppressed, too.

Depending on the camera's capturing mode, it is also necessary to transform provided radial distances to cartesian distances, using (11):

$$X = \frac{df}{\sqrt{u^2 + v^2 + f^2}} \quad (3)$$

## 4.2 Construction of U-V-Histograms

**Construction.** The calculation of u-v-histogram images follows the approach presented in [9]. Taking the pitch angle  $\theta$  into account, the projection model from (11) can be described by:

$$\begin{aligned} u &= f \frac{Y}{Z \sin \theta + X \cos \theta} + p_x \\ v &= f \frac{Z \cos \theta - X \sin \theta}{Z \sin \theta + X \cos \theta} + p_y \end{aligned} \quad (4)$$

Every plane being parallel to the Y axis is defined through:

$$aX + bZ = c, \quad (5)$$

where  $a, b, c$  denote constant variables, depending on  $\theta$ . Combining with equation (4), the following linear equation can be deduced:

$$v = f \left( \frac{c}{bX} - \frac{a}{b} \right) + p_y \quad (6)$$

This can be expressed simplified by:

$$v = f \frac{1}{X} + p_y \quad (7)$$

Equation (7) shows that all planes parallel to the Y axis can be projected as a straight line in dependency of the reciprocal depth into the v-histogram.

Similar, planes parallel to the Z axis are defined through:

$$aX + bY = c \quad (8)$$

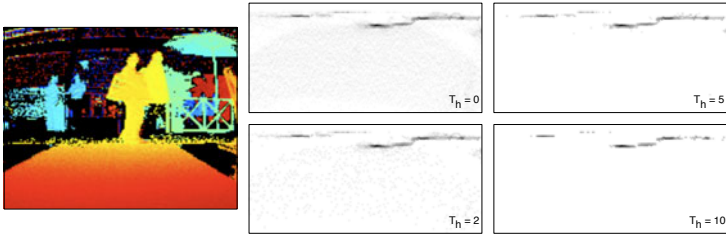
The simplified equation in combination of (4) is:

$$u = f \frac{1}{Z \sin \theta + X \cos \theta} + p_x \quad (9)$$

Considering the pitch angle  $\theta$  to be small, all planes parallel to the Z axis can be projected as straight lines in dependency of the reciprocal depth into the u-histogram.

The dimension of the u-(v-)histogram image is defined by the number of columns (rows) of the input image and the number of histogram bins.

Results of the u-v-histogram calculation are illustrated in Fig. 3. Without reciprocal distance correction, the planes parallel to the Y axis cannot be projected as straight lines (see Fig. 3 right).



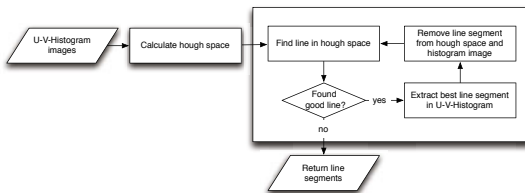
**Fig. 4.** Noise reduction of u-histogram using accumulator threshold  $T_h$

**Noise reduction.** The constructed histogram images show distinct lines where obstacle planes are parallel to the Z axis (u-histogram) respectively Y axis (v-histogram). However, planes not normal to the projection plane lead to noise in the histogram images. Therefore, we analyze the histogram distribution and filter points having an accumulator value smaller than a predefined threshold  $T_h$ . Fig. 4 shows the resulting u-histogram with threshold values  $T_h = \{0, 2, 5, 10\}$ .

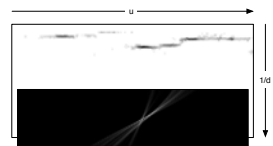
### 4.3 Line Segment Extraction

Fig. 5 depicts the flowchart of the line segment extraction. The algorithm starts transforming the u-(v-) histogram image into its corresponding Hough space. In the Hough space, the highest value defines the best line in the histogram image. The best segment of the line is searched using a modified line pattern analysis. Afterwards, the line segment points are removed from both the Hough space and the histogram image. The algorithm continues analyzing the next best value in the Hough space, until a certain abort criteria is reached. This can either be a maximum number of line segments already extracted or no more distinct points in the Hough space.

**Hough Transformation.** After histogram construction, straight lines must be extracted in both u- and v-histogram images. Thereby we use a modified Hough transformation. By definition, no vertical lines can exist in the u-histogram image. So we are able to narrow the Hough space's dimension to search for straight



**Fig. 5.** Flowchart of the line segment extraction algorithm



**Fig. 6.** Resulting Hough space of u-histogram image

lines only for angular values  $\omega = [0 + \epsilon, \pi - \epsilon]$ , where  $\epsilon$  is given by the minimum reflection angle before total illumination light suffers of total reflection. Likewise, no horizontal lines can exist in the v-histogram image, so the dimension of the Hough space is limited to angular values  $\omega = [-\frac{\pi}{2} + \epsilon, \frac{\pi}{2} - \epsilon]$ . A resulting Hough space is shown in Fig. 6.

Because the dimension of the u-v-histograms is defined by the input image size and  $bin_n$ , the Hough space can be pre calculated to reduce the computations at runtime.

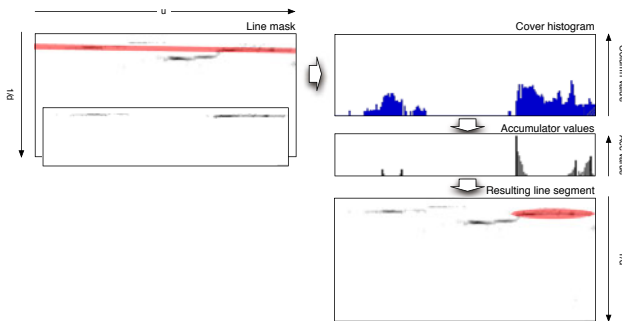
Instead of convolution of the image data with a filter mask, the histogram values can be projected directly into the Hough space. Besides, the Hough data is weighted by the histogram’s accumulator value, such as distinct points in the histogram cause higher peaks in the Hough space:

$$r_u(\omega) = \left( u \cos \omega + \frac{1}{d} \sin \omega \right) \text{acc} \left[ u, \frac{1}{d} \right] \tag{10}$$

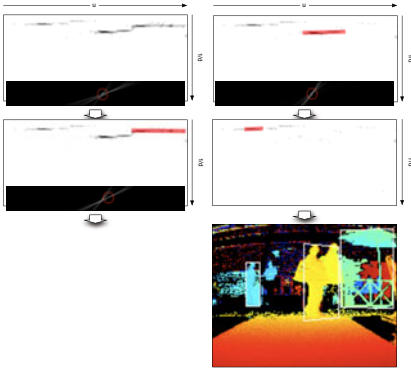
$$r_v(\omega) = \left( \frac{1}{d} \cos \omega + v \sin \omega \right) \text{acc} \left[ \frac{1}{d}, v \right] \tag{11}$$

**Line pattern analysis.** Maximum values in the calculated Hough spaces correspond to strong lines in the histogram images. By definition, lines given by Hough parameters  $(r, \omega)$  do not have start or end points. Hence, the next step is to extract line segments. For this we calculate a cover histogram, considering only the points in the u-(v-)histogram image matched by the line defined by  $(r, \omega)$ , and some  $\epsilon$  region (see Fig. 7).

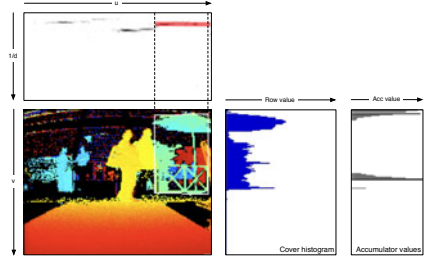
All histogram values defined by this mask are accumulated per column (row) in the u-(v-)histogram image, so that each column (row) along the line has its own cover value. Next, we apply a line pattern voting algorithm. Here, we use a modified approach to the well known LPHT presented by Yip [11]. The original LPHT searches for start and end points in a 2D image by accumulating



**Fig. 7.** Line pattern analysis. (1<sup>st</sup> image) Line mask with  $\epsilon$  region. (2<sup>nd</sup> image) Cover histogram of line mask. (3<sup>rd</sup> image) Accumulated values from line pattern voting algorithm. (4<sup>th</sup> image) Resulting line segment.



**Fig. 8.** Subsequently removing extracted line segments from Hough image. After 3rd segment extraction no more distinct lines are found.



**Fig. 9.** Line pattern analysis for vertical dilation. (1<sup>st</sup> row) Extracted line segment. (2<sup>nd</sup> row, left image) Resulting plane. (2<sup>nd</sup> row, middle image) Cover histogram of depth values. (2<sup>nd</sup> row, right image) Accumulated values from line pattern voting algorithm.

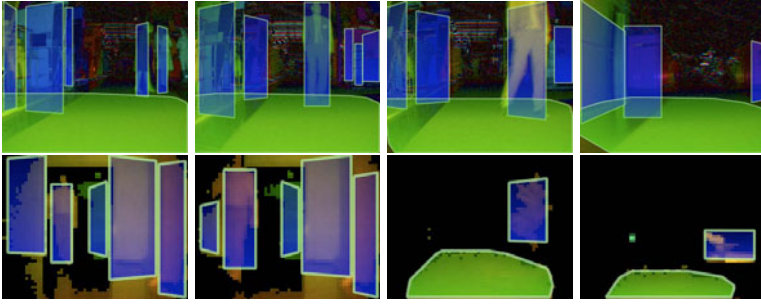
relative connectivity for each point in the image. In contrast, we calculate relative connectivity only along the 1D line histogram (see Fig. 7), which speeds up computation time from  $O(MN \log(M) \log(N))$  for the LPHT to  $O(N \log(N))$  for our method. Details of the algorithm are further explained in [12].

After line segment extraction, the points belonging to the segment are removed from both the Hough space and the histogram image. Removing the points from the Hough space and histogram image allows the segmentation method to be sensible for small segments as well. Thus, the algorithm is able to find all line segments in the histogram images (see Fig. 8).

Each line segment in the u-(v-)histogram is then defined with its image column (row) start and end point and the corresponding depth indices of the start and end points. During histogram calculation, the reciprocal depth values were sorted into histogram bins, so the depth indices must be mapped to the corresponding depth range in the depth image. Yet, the height (width) of the extracted u-(v-)segments in image rows (columns) is unknown, hence the segments must be analyzed by subsequent algorithms.

#### 4.4 Object Detection

For object detection, we consider only line segments extracted in the u-histogram image. For each line segment, the object height can be reconstructed using the original depth image. First we select all points in the depth image, which lie between start and end point of the line segment. Next, we calculate a cover histogram along the image rows. Therefore, we subsequently check all image rows and accumulate the histogram's row value, if a point's depth value lies in the depth range of the line segment. Then, we check relative connectivity of the cover histogram again using the 1D-LPHT explained above to find the line



**Fig. 10.** Experimental results of the proposed segmentation algorithm using SR 4000 (upper row) and PMD 3kS2 (lower row) tof cameras on a moving platform

segment along the  $v$  image axis (see Fig. 9). The resulting object plane is then given by start and end points of the histogram line segment and the vertical line segment.

#### 4.5 Floor Detection

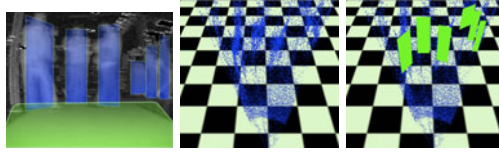
To find all points belonging to the floor plane, the algorithm must differentiate between object and floor plane detection. During object detection, all planes parallel to the  $Z$  axis should be extracted. For floor plane detection, only those planes which lie in the  $X$ - $Y$  plane should be considered, with regards to the pitch angle  $\theta$  of the camera to the  $X$ - $Y$  plane. Therefore, we compare all line segments in the  $v$ -histogram with Hough gradient  $\omega$  to the expected floor angle in terms of segment length and angular offset. The best line segment is then considered as the floor plane.

To find all image points belonging to the floor plane, image points in the depth image are compared with the floor line segment. Again, we select all points in the depth image, which lie between start and end point of the line segment. If the points' depth value lie in the depth range of the histogram line segment, they are added to the set of floor points. The resulting floor plane is computed as the convex hull of all points using Grahams scan algorithm [13]. Fig. 10 shows segmentation results with the floor plane in green color.

## 5 Experiments

### 5.1 Sensor Setup

Our sensor setup consists of one standard video camera and two tof cameras. The near range tof camera is a SR 4000 with a resolution of  $176 \times 144$  pixels and an unambiguity interval of  $15m$ . The long range camera is a PMD 3kS2 with a much smaller resolution of  $64 \times 50$  pixels but an increased unambiguity interval of  $120m$ . The illumination units of the 3kS2 provide enough light for about  $40m$  even in outdoor environments. For more details on the hardware platform see [14].



**Fig. 11.** Experimental results (left), SR 4000 point cloud (middle), point cloud with resulting object planes (right)

## 5.2 Experiments

We have evaluated the proposed segmentation algorithm with real data under varying conditions acquired with the SR 4000 tof camera as well as the PMD 3kS2 tof camera. As no ground truth information is available, a qualitative performance evaluation is conducted.

The set of images displayed in the upper row of Fig. 10 show the result of our proposed algorithm using the SR 4000 tof camera images. The images were captured during CeBit 2009 exhibition. The planes displayed in blue are object planes, the green plane shows the segmented floor plane.

The algorithm is able to segment all distinct planes, even those almost parallel to the driving direction of the camera platform ( $4^{th}$  image). Over-segmentation only happens near to the floor plane, where light is reflected from objects by the partly specular floor plane. Under-segmentation however can occur rarely, mostly at object ends. The detected floor plane can be recognized robustly.

Another set of images shows segmentation results using the PMD 3kS2 tof camera (see Fig. 10 lower row). The first two images were captured indoor, the last two in outdoor environment. As the resolution of the PMD camera is much smaller, segmentation of object and floor planes must be performed with fewer plane points. The algorithm still works well, but object edges are under-segmented more frequently. The floor detection fails in the first two images, caused by a highly reflecting floor. The floor detection works fine on paved ground, as seen in the last two images.

With both cameras, segments are split in some frames, commonly when lines in the histogram images are noisy. However, these segments can be remerged using temporal fusion. Our proposed method does segment distinct objects, no matter if they are parallel to the image plane or not. As mentioned in Section 2, this is a drawback of depth based segmentation methods commonly used for tof data segmentation.

## 5.3 Computation Time

The computation time needed for segmentation depends strongly on the image resolution and the number of valid pixels recorded by the camera. Thus, the amount of time needed for image processing is different for the SR 4000 tof camera compared to the PMD 3kS2 tof camera. For speed comparison, the proposed method was applied to a scene where both cameras had few invalid pixels, the

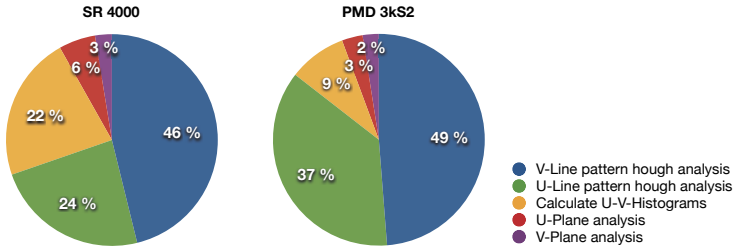


Fig. 12. Percentage of calculation time per module

algorithm was run single threaded on a standard desktop computer. The computation time for the SR 4000 tof camera takes approximately  $0.04s - 0.06s$ , while the PMD 3kS2 tof camera takes about  $0.02s - 0.03s$  per frame. Fig. 12 shows the percentage of the calculation time needed by the most relevant modules.

## 6 Conclusions and Future Works

We presented a model free segmentation approach that is capable of segmenting object and floor planes in unstructured environments. The algorithm is using statistical knowledge by accumulating u-v-histogram images for initial plane detection. Hough transformation is used for line extraction, while the concept of line pattern voting provides a fast approach for line segment detection. Line segments are then analyzed to reconstruct the full dimension of the segmented 3D structure. Although the segmentation uses projection planes for robust detection, no information is lost, as the full 3D structure is considered during segmentation. We demonstrated that the proposed algorithm achieves good results on data acquired in unstructured environments. Further steps include the automatic generation of drivable paths around detected objects upon the floor plane.

## Acknowledgments

This work was partially supported by the Ministry of Economic Affairs of Baden-Württemberg.

## References

1. Thrun, S.: Learning occupancy grid maps with forward sensor models. *Autonomous Robots* 15, 111–127 (2003)
2. Skuttek, M., Linzmeier, D., Appenrodt, N., Wanielik, G.: A precrash system based on sensor data fusion of laser scanner and short range radars. In: *8th International Conference on Information Fusion 2005*, vol. 2 (July 2005)



3. Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P.J., Bunke, H., Goldgof, D., Bowyer, K., Eggert, D., Fitzgibbon, A., Fisher, R.: An experimental comparison of range image segmentation algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* 18, 673–689 (1996)
4. Moosmann, F., Pink, O., Stiller, C.: Segmentation of 3d lidar data in non-flat urban environments using a local convexity criterion. In: *2009 IEEE Intelligent Vehicles Symposium*, pp. 215–220 (2009)
5. Keller, M., Kolb, A.: Real-time simulation of time-of-flight sensors. *Simulation Modelling Practice and Theory* 17, 967–978 (2009)
6. Fardi, B., Dousa, J., Gerd, W., Elias, B., Barke, A.: Obstacle detection and pedestrian recognition using a 3D PMD camera. In: *Proceedings of Intelligent Vehicles Symposium*, pp. 225–230. IEEE, Los Alamitos (2006)
7. Parvizi, E., Wu, Q.: Real-time approach for adaptive object segmentation in time-of-flight sensors. In: *20th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2008*, vol. 1, pp. 236–240 (November 2008)
8. Schamm, T., Zöllner, J.M., Vacek, S., Schröder, J., Dillmann, R.: Obstacle detection with a photonic mixing device-camera in autonomous vehicles. *International Journal of Intelligent Systems Technologies and Applications* 5, 315–324 (2008)
9. Labayrade, R., Aubert, D., Tarel, J.-P.: Real time obstacle detection in stereovision on non at road geometry through “v-disparity” representation. In: *Intelligent Vehicle Symposium*, vol. 2, pp. 646–651. IEEE, Los Alamitos (2002)
10. Hu, Z., Lamosa, F., Uchimura, K.: A complete u-v-disparity study for stereovision based 3d driving environment analysis. In: *3-D Digital Imaging and Modeling*, pp. 204–211 (June 2005)
11. Yip, R.: Line patterns hough transform for line segment detection. In: *Proceedings of 1994 IEEE Region 10’s Ninth Annual International Conference on Theme: Frontiers of Computer Technology, TENCON 1994*, vol. 1, pp. 319–323 (1994)
12. Teutsch, M.: Fusion von 6D-Stereo- und Radardaten zur Segmentierung und Attributierung von Objekten im Straßenverkehr. Diploma’s thesis, FZI Forschungszentrum Informatik (January 2009)
13. Graham, R.: An Efficient Algorithm for Determining the Convex Hull of a Finite Planar Set. *Information Processing Letters* 1, 132–133 (1972)
14. Schamm, T., Strand, M., Gump, T., Kohlhaas, R., Zöllner, J., Dillmann, R.: Vision and ToF-based driving assistance for a personal transporter. In: *International Conference on Advanced Robotics, ICAR 2009* (June 2009)

# Scale Invariant Feature Transform (SIFT) Parametric Optimization Using Taguchi Design of Experiments

Dominic R. Maestas, Ron Lumia, Gregory Starr, and John Wood

Department of Mechanical Engineering, University of New Mexico,  
New Mexico USA, 87110  
maestasd@unm.edu

**Abstract.** Traditional SIFT methods require a priori of object knowledge in order to complete accurate feature matching. The usual means is via trained databases of objects. In order to be able to get the pose of an object, accurate object recognition is required. Without accurate object recognition, detection can occur but no information about 3-D location will be available. The goal of this work is to improve object recognition using SIFT by optimizing algorithm parameters with respect to the mean angle between matched points ( $\mu_{AMP}$ ) found within a scene via multiple images, which can then be used to determine the object pose. Good parameters are needed so that the SIFT algorithm is able to control which matches are accepted and rejected. If keypoint information about an object is wrongly accepted, pose estimation is inaccurate and manipulation capabilities in a 3-D work space will be inaccurate. Using optimized SIFT parameter values results in a 19% improvement of  $\mu_{AMP-Optimal}$  in comparison to  $\mu_{AMP-Experimental}$ .

**Keywords:** Object Recognition, Scale Invariant Feature Transform, Taguchi, Design of Experiments.

## 1 Introduction

Improving SIFT algorithm performance by choosing the best set of parameters is a challenge that remains unresolved in the field of object recognition. In order to perform object recognition using SIFT, it is essential that its key parameters such as, octaves levels, the number of times the Difference of Gaussians is completed, and its thresholds are sufficiently set so that under the most constrained conditions feature extraction from consecutive images is consistent. The first constraint assumes that lighting conditions, such that illumination, are constant over the field of view. Secondly, adequate matching must be possible on a single object from multiple images from a single camera or single stereo view. Lastly, once single camera optimization occurs, multiple camera views can be integrated into an algorithm that estimates pose.

SIFT matching and keypoint detection has proved to be effective in overcoming object occlusions during object recognition. The main purpose of the following experiments is to optimize SIFT parameters in order improve object recognition capabilities that can be used to estimate the pose of an object.

In order to optimize the SIFT functional framework parameters, a full factorial experiment could be undertaken to test every possible combination of parameters to choose those parameters that ensure proper matching, i.e., the best correspondence between feature points on two images. Alternatively, Taguchi Design of Experiments (DOE) will be the method of choice since its experiments are able to reveal the effects of each parameter and can be optimized within the algorithm. L8 design is best suited since there are 6 key parameters that control the thresholds of keypoints, frames, and matches between comparative images.

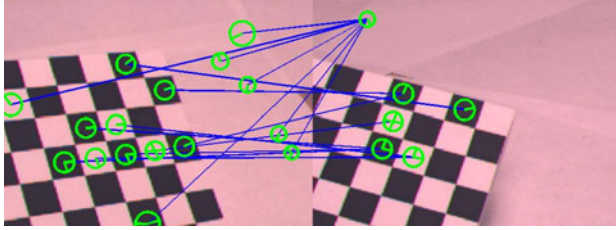
Using a histogram approach provides insight into obvious mismatching errors. Keypoints that are different in comparative images, but matched using SIFT, will have an angle between the two corresponding points that deviates from the rest of the angles between matches. In the histogram, these mismatches will appear as outliers and prime candidates for elimination during object recognition by optimizing SIFT parameters. It is expected that in a perfect match situation, the set of matched feature points from one image to the other form parallel lines.

## 2 Previous Work

Algorithms that make use of Scale Invariant Feature Transform (SIFT) and Maximally Stable Extremal Region (MSER) extract information using vision systems [1,4,5,6]. Given a database trained with keypoint and descriptor information of an object, it is possible to categorize features for object recognition[8, 11]. However, as an object is manipulated in a 3-D workspace, the ability to recognize an object is based on real-time image matching and then estimating the pose. This new approach can leverage SIFT framework modified to create a new matching algorithm that uses invariant keypoint features based on four corresponding views of an object.

### 2.1 Scale Invariant Feature Transform (SIFT)

One of the approaches in extracting image information was introduced by David Lowe and is known as the Scale Invariant Feature Transform (SIFT). Its primary purpose was image matching in order to gain information about structure for object recognition. To use the SIFT algorithm, one choose a number of parameters, e.g., octaves, levels, and thresholds. Lowe used local reference frames to show that scale and rotation do not change from pixel to pixel [5,6]. This property is referred to as invariance in the local frame. SIFT works because of the large number of keypoints it generates from an image [5]. The density of keypoints allows for frames to be created in clusters. Small objects in a scene background can be recognized as long as 3 features are correctly matched, either from two images or from a training database via Euclidean distance [5]. An example of SIFT matching is shown in figure 1 below. In each checkerboard square, a keypoint is established based on the difference in contrast to its nearest neighbor. Then, the scale is established and indicated by the size of the circle. Next, the orientation of the object is shown using the tick mark within the circle to identify its angular position. Finally, the features are compared to one another between two images and lines reveal which points SIFT computes as matches.



**Fig. 1.** SIFT Matching

Algorithms are written to take advantage of the large number of features that densely fill the image scene based on Lowe's four stages of image matching: scale-space maxima detection, keypoint localization, orientation, and keypoint descriptors [5, 6].

Cheung successfully demonstrated SIFT object recognition by matching medical image scans against a trained database [2]. Grabner, et al. were able to improve the computational efficiency of feature matching in terms of faster processing [3] and provided a framework that can be used to recover object recognition information using parametric analysis.

Vedaldi and Fulkerson used Lowe's SIFT concept as the basis of generating generic functions that could be used in algorithms to create a matrix containing  $x$ ,  $y$ , scale, and orientation features [10].

VLFeat is an open source library that uses Lowe's SIFT algorithm to create functions that can be used in matching algorithms, such as the one proposed below, in order to extract frames. It takes into account illumination and workspace viewpoints in order to create descriptors and keypoints used for matching [10].

## 2.2 Taguchi Design of Experiments (DOE)

Design of Experiments (DOE) is an approach that computes variation in a process in order to minimize the number of experiments required to achieve an optimal set of parameters. Traditionally, DOE has been used to determine the optimal setting for machine parameter values. In our research, DOE is used to determine the optimal values for the parameters of the SIFT algorithm.

The Taguchi method, a particular type of DOE algorithm is a way to plan, conduct, analyze, and determine optimal settings using orthogonal arrays. Orthogonal arrays are able to show distinction between control factors and noise factors. Furthermore, the orthogonal array is the method by which relatively few experiments span a large experiment space [9]. The major advantage of the Taguchi method, similar to all DOE methods, is that the use of orthogonal arrays allows the user to perform far fewer experiments than would be required from a full factorial set of experiments.

Since the Taguchi DOE method performs relatively few experiments, it is unlikely for the optimal answer to be one of the prescribed experiments. However, it is possible to use the experiments to determine the optimal settings as well as to predict the output expected from the use of those settings.

### 3 Object Recognition Concept

The basis for this research uses SIFT functions to create a framework that accurately detects and extracts information which is used for object recognition. The research focus is on extracting features from an object located on a vision table with 4 cameras located at each corner in order to perform object recognition and use the information to determine its pose.

SIFT is used to find keypoints and frames in image scenes in the following object recognition algorithm. Using SIFT, the algorithm recognizes an object via feature extraction. Then the pose can be determined and fed into a robot controller to direct a whole arm manipulator (WAM) used to perform manipulation tasks.

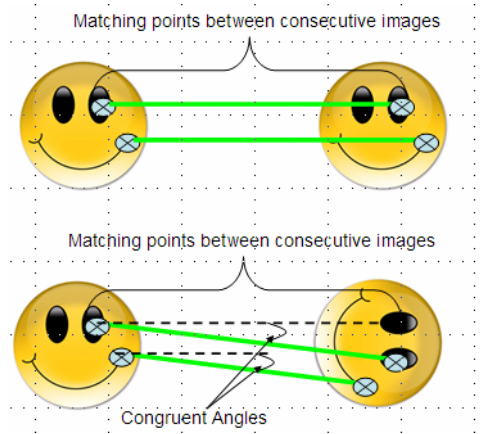
The actual optimization occurs on SIFT parameters: octaves, levels, and thresholds. The parameters use the Difference of Gaussians to control object recognition.

The mean angle between matched points ( $\mu_{AMP}$ ) of consecutive images is the index used as the Taguchi measure of goodness. It is calculated using the following equation:

$$\mu_{AMP} = \frac{\sum_{i=1}^n AMP_i}{n}, \text{ where } AMP_i \text{ is the angle between each pair of keypoints, and } n$$

is the number of matches.

It is a good measure because matching points on the same object will have the same angle (e.g. parallel lines) as shown in figure 2 below.



**Fig. 2.** Example of Matching Points and Their Angles

The angles provide a visual cue that indicates which points can be excluded from the  $\mu_{AMP}$  calculation. Ideally, all matched feature points will form parallel lines, as shown in figure 2. The  $\mu_{AMP}$  will improve by better choice of SIFT parameters. The  $\mu_{AMP}$  determines which pairs of matching points to utilize for pose estimation during object recognition.

Optimizing SIFT parameters is achieved through experiments using Taguchi DOE. The experiment is setup up by the number of factors we are interested in controlling. Since there are 6 parameters, a Taguchi L8 design matrix is sufficient enough to determine the effects and interactions between the orthogonal arrays.

### 3.1 Image Detection Implementation

The basis for our application is looking at an object located at the center of the vision table with multiple cameras under different fields of view. The algorithm capable of object recognition is composed of the following six steps.

1. Using the 4 camera system, determine which two image sources provide the best view of the object.
2. Capture images of the object under consideration from the two chosen cameras.
3. Extract the position information using SIFT frames for each image.
4. Determine the angle between SIFT points using corresponding frames from each image.
5. Determine the geometric mean based off the histogram of angles between SIFT points.
6. Eliminate undesired matches via Statistical Based Limits.

Within the SIFT code, different values of parameters for the octaves, levels, and thresholds, produce different qualities of matching results. Taguchi DOE is used to find the set of SIFT algorithm parameters that produces the best matching.

The goal of the experiments is to determine the set of SIFT algorithm parameters that creates matches that produce the most parallel lines. (See figure 2.) The  $\mu_{AMP}$  is the index of performance that is used to optimize the parameter set.

## 4 Experimentation

### 4.1 Experimental Setup

The equipment used during the experiments includes 4 Basler A605fc-2 cameras assembled on a vision table with 1 camera mounted via Panavise 15" goosenecks at each corner, lenses were Fujinon HF9HA-1B with a 1:1.4/9mm lens. Lighting is controlled using two umbrella light assemblies to ensure diffuse illumination capable of 250-650 watts. The vision table is connected via IEEE 1394 fire wires to two PCI-fire wire cards (2 cameras per card) as shown in Figure 3 below. The station controller is run with a core 2 duo processor from Intel, with 3 gigabytes of onboard Random Access Memory (RAM). Experimental analysis used Matlab with Signal Processing, Video and Image Processing, and Image Acquisition tool boxes during algorithm implementation. JMP 7.0 was the basis for the Taguchi Design, effects analysis rollup, and optimal parametric prediction.

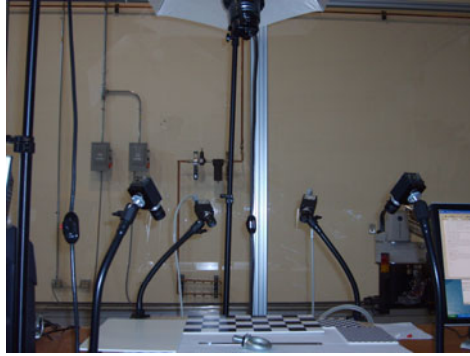


Fig. 3. Experiment Setup with Vision Table

## 4.2 Parameters

The following list describes the six key parameters for the SIFT algorithm. In addition, the low/high values of each variable used by the Taguchi DOE are shown.

1. **Octaves: (low=2 high=8)** is the number of times that the Difference of Gaussians is performed in order to create a scale space.
2. **Levels: (low=3 high=4)** determine the number of levels per octave within the Difference of Gaussians subspace.
3. **First Octave: (low=0 high=1)** is where to start in the Difference of Gaussian.
4. **Peak Threshold: (low=0 high=1)** is the maximum value that the algorithm will accept during matching.
5. **Edge Threshold: (low=10 high=20)** determines the threshold of non-edge keypoint selection within the image structure.
6. **Normal Threshold: (low=2 high=8)** is the minimum 12-norm of a descriptor before it is normalized and below this threshold is railed to zero.

## 4.3 Taguchi Setup

The experimental set up consists of 8 conditions tested over two runs for high and low changes in the parameters using the L8 design codes, shown below, where (–) is the low value of the parameter and (+) is the high value of the parameter.

1. – – – – –
2. – – – + + +
3. – + + – – +
4. – + + + + –
5. + – + – + –
6. + – + + – +
7. + + – – + +
8. + + – + – –

L8 code for DOE 1 corresponds to the coded pattern (-----) which tests the low values of each parameter. The parameter settings are: Octaves=2, Levels=3, First Octave=0, Peak Threshold=0, Edge Threshold=10, and Normal Threshold=2 executed over two separate random runs and then averaged together using the following equation:

$$\frac{\sum x}{n} \tag{1}$$

The signal to noise ratio is computed with:

$$-10 \log_{10} \left( \frac{\sum \frac{1}{x^2}}{n} \right) \tag{2}$$

where x is the value of the mean between angles of the keypoints, and n is the number of runs.

**Table 1.** L8 Design for SIFT Parameters

1	Octaves	Levels	FirstOctave	PeakThresh	EdgeThresh	NormThresh	Pattern
2	2	3	0	0	10	2	-----
3	2	3	0	1	20	8	-----+
4	2	4	1	0	10	8	-----+
5	2	4	1	1	20	2	-----+
6	8	3	1	0	20	2	++---+
7	8	3	1	1	10	8	++---+
8	8	4	0	0	20	8	++---+
9	8	4	0	1	10	2	++---+

Once the mean and the signal to noise ratios are calculated, the effect of each factor is known and optimized to determine the best possible combination of parameters used when completing feature matching. The result leads to the greatest possible number of matches and reduces the number of invalid matching between comparative images.

To compute the optimal values,  $\eta$ , Taguchi’s Additive Model (no interaction of parameters is assumed) for orthogonal arrays [7] is used:

$$\eta = \mu + \sum_1^8 (\bar{x}_j - \mu) \tag{3}$$

where  $\mu$  is the overall mean,  $x_j$  is the effect mean from 1 to j, the number of experiments.

#### 4.4 Results

Using a random run order, 16 experiments (8 experiments across 2 runs) were conducted and the results are shown in Table 2 corresponding to the L8 design codes



**Table 2.** L8 Results for SIFT Parameters

1	Octaves	Levels	First Octave	PeakThresh	EdgeThresh	NormThresh	Mean
2	2	3	0	0	10	2	9.14
3	2	3	0	1	20	8	0
4	2	4	1	0	10	8	-2.58
5	2	4	1	1	20	2	0
6	8	3	1	0	20	2	-8.20
7	8	3	1	1	10	8	0
8	8	4	0	0	20	8	-10.54
9	8	4	0	1	10	2	0

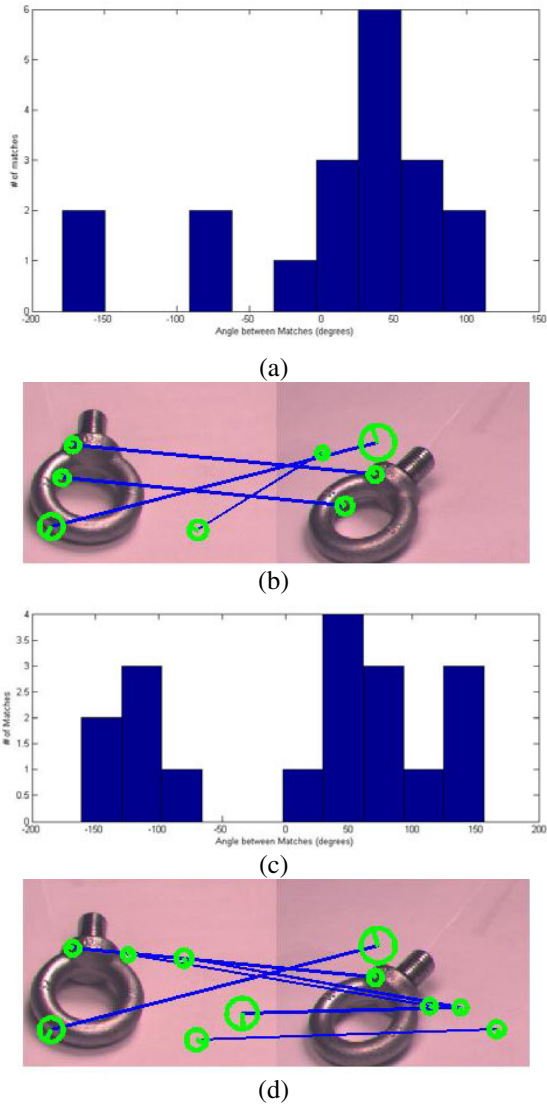
(octaves, levels, first octave, peak threshold, edge threshold, norm threshold) and shown in the table below. Using equations (1) and (2) the mean and signal to noise ratio were calculated for each run in order to estimate the optimal parameters to use for feature matching.

#### 4.5 Analysis

The  $\mu_{AMP-Experimental}$  for the subset of experiments is  $-1.52^\circ$ . The first analysis revealed that one of main effects due to the signal factor of peak threshold always lead to a mean of zero when it was at a level of 1. Because of this effect it provides, as Taguchi calls it, a way to block. Blocking identifies factors that may not be a relevant source of variation and can be set to minimize the factor's impact to the algorithm. As a result, runs associated with a signal level of 1 eliminated matching and resulted in a mean of zero due to the results being blocked out. The output of our algorithm confirms this as shown in figure 4 below since the threshold allows for absolute control of what matching is accepted—in this case none. Follow up experiments verified that this effect, down to a signal level of 0.25, rejected all matches. At levels less than 0.25 the results were insignificant with respect to a level set to zero.

**Fig. 4.** Matching Output at Peak Threshold

Experiment 1 was conducted using all the settings (octaves, levels, first octave, peak threshold, edge threshold, norm threshold) at their low values (2,3,0,0,10,2). The resulting angle measurement had a mean of  $9.14^\circ$ , with a signal to noise ratio of 19.18 dB. The histogram and matching results are shown in figure 5a-5d below.



**Fig. 5.** (a) Run 1 Histogram (b) Run 1 Matching Output (c) Run 2 Histogram (d) Run 2 Matching Output

Experiment 3 ran with the following parameter values (2,4,1,0,10,8) and showed convergence to zero in its mean of  $2.58^\circ$  and a signal to noise ratio of 8.22 dB. Experiment 5 ran with the following parameter values (8,3,1,0,20,2) and resulted in a mean of  $-8.20^\circ$  and a signal to noise ratio of 13.59 dB. Experiment 7 ran with the following parameter values (8,4,0,0,20,8) and resulted in a mean of  $-10.54^\circ$  and a signal to noise ratio of 17.34 dB.

From the histogram, it is seen that the effect on the mean is largely influenced by outliers. The figure also shows that based on the thresholds under half of the matches are accepted and do not provide positive matching results that are useful in recovering object information. In experiment 3, this is explained because of where the first octave starts, at its high value, and matching from consecutive images converges on to a single point of the second image. Object recognition is achieved but no valuable information is able to be attained to determine pose as a result of the negative matches. In experiment 5 the first octave starts the same as experiment 3--its high value. Some positive matching seems to be occurring. As far as object recognition and pose determination, limited to no information is attained as a result of the number of negative matches. Experiment 7 had a dense number of keypoints and descriptors. Matching between the images appears to converge as SIFT predicts but no object recognition information is able to be attained as a result of the improper convergence.

#### 4.6 Optimal Prediction

The goal of Taguchi DOE is to run a set of experiments using orthogonal arrays as opposed to running a full factorial design, in this case  $2^8$  experiments. Then, a prediction of the optimal parameters is made using Taguchi's Additive Model so that positive matching leads to improved object recognition.

The  $\mu_{AMP-Experimental}$  is  $-1.52^\circ$ . The next step is to investigate how the effects and levels are impacted by each factor. The information from the effects roll up, as shown in figure 6, and equation (3) are the basis of determining the optimal parameters based on the L8 results aforementioned.

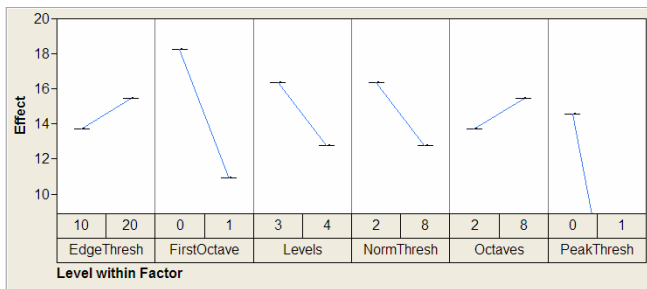


Fig. 6. Effects Chart for Each Parameter

From figure 6, we can see how each parameter affects the object recognition algorithm. Then we can choose the parameter setting with the highest effect to improve the algorithm matching. Choosing the highest effect on each reveals the following optimal parameters:

1. **Octaves** = 8
2. **Levels** = 3
3. **First Octave** = 0
4. **Peak Threshold** = 0
5. **Edge Threshold** = 20
6. **Normal Threshold** = 2

Lastly, the experiment is repeated for the optimal setup. The  $\mu_{AMP-Optimal}$  converges toward zero and is equal to  $-1.278^\circ$  as shown in the histogram below. It is observed that the distribution is normal between  $-50^\circ$  to  $50^\circ$  and is consistent with SIFT matching as shown in figure 7 below. Statistical based limits can be implemented to control the threshold to reject angles outside this range and reduce the number of false matches associated with the algorithm. The result is a 19% improvement of the optimal  $\mu_{AMP-Optimal}$  in comparison to non-optimized experimental  $\mu_{AMP-Experimental}$ .

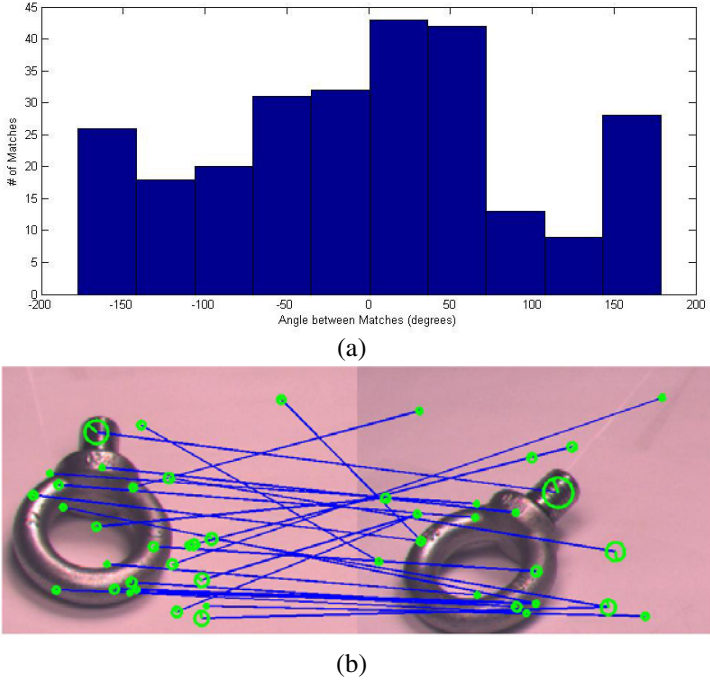


Fig. 7. (a) Run 1 Histogram (b) Run 1 Matching Output

## 5 Conclusion

Taguchi DOE was used to choose the best set of SIFT algorithm parameters so as to improve object feature matching between two views of the same object.

In the future, we will use the feature matches to identify objects using a database with features stored a priori. Furthermore, we will compute object pose from the matches through camera calibration techniques. Finally, we intend to develop algorithms that will automatically decide which of the four camera images are occluded by other objects, e.g., a robot gripper carrying another object.

**Acknowledgments.** Thanks to David Vick and Colin Selleck for help in algorithm code development. This work was supported in part by the DOE URPR Grant # DE-FG52-04NA25590.

## References

- [1] Cao, F., Morel, J., Muse, P.: A Theory of Shape Identification. Lecture Notes in Mathematics, vol. 1948. Springer, Heidelberg (2008)
- [2] Cheung, W.: n-SIFT: n-dimensional scale invariant feature transform for matching medical images. In: Proceedings of the Fourth IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 720–723 (2007)
- [3] Grabner, M., Grabner, H., Bischof, H.: Fast approximated SIFT. In: 7th Asian Conference of Computer Vision, pp. 918–927 (2006)
- [4] Kalpakjian, S., Schmid, S.: Manufacturing Engineering and Technology, 5th edn., pp. 1144–1189. Pearson Prentice Hall, London (2006)
- [5] Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91–110 (2004)
- [6] Lowe, D.: Object Recognition from Local Scale-Invariant Features. In: IEEE International Conference on Computer Vision, vol. 2, pp. 1150–1157 (1999)
- [7] Lumia, R.: ME 586: L11 Taguchi Design of Experiments, pp. 1–13. University of New Mexico (2008)
- [8] Montesano, L., Lopes, M., Bernardino, A., Santos-Victor, J.: Learning Object Affordances: From Sensory–Motor Coordination to Imitation. IEEE Transactions on Robotics 24(1) (2008)
- [9] Taguchi, G., Chowdhury, S., Wu, Y.: Taguchi’s Quality Engineering Handbook, 1696 pages. John Wiley & Sons Inc., Chichester (2005)
- [10] Vedaldi, A., Fulkerson, B.: VLFeat: An Open and Portable Library of Computer Vision Algorithms (2008), <http://www.vlfeat.org/>
- [11] Zhang, W., Yang, H., Samarasinghe, D., Zelinsky, G.: A Computational Model of Eye Movements during Object Class Detection. In: Advances in Neural Information Processing (NIPS), Vancouver, Canada, vol. 18, pp. 1609–1616 (2005)

# Robust Visual Monitoring of Machine Condition with Sparse Coding and Self-Organizing Map

Haining Liu, Yanming Li, Nan Li, and Chengliang Liu

School of Mechanical Engineering, Shanghai Jiao Tong University,  
Shanghai 200240, China

**Abstract.** A direct way to recognize the machine condition is to map the monitored data into a machine condition space. In this paper, via combining Sparse Coding and Self-Organizing Map, a new model (SC-SOM) is proposed for robust visual monitoring of machine condition. Following the model, a Machine Condition Map (MCM) representing the machine condition space is formulated offline with the historical signals; then, during the online monitoring, the machine condition can be determined by mapping the monitoring signals onto the MCM. The application of the SC-SOM model for bearing condition monitoring verifies that the bearing condition can be correctly determined even with some disturbances. Furthermore, novel bearing conditions can also be detected with this model.

**Keywords:** Fault diagnosis, Condition monitoring, Sparse coding, Neural networks, Self-organizing maps.

## 1 Introduction

Condition-based maintenance (CBM) has emerged as an effective strategy to prevent engineering systems from catastrophic failures while minimizing maintenance cost. Proper maintenance decision can only be made based on correct recognition of the machine condition. A direct way to achieve this is to map the monitored data into a machine condition space [1,2]. Generally, the construction of the machine condition space follows such a roadmap: data acquisition, feature extraction and dimension reduction. Feature extraction is essential due to the fact that the acquired raw data are always in high dimension and the amount of data is massive. Numerous features need to be extracted to construct the machine condition space, while the dimension reduction makes the distribution of each subspace representing one machine condition more explicit. A well-defined machine condition space should contain subspaces as more as possible to encompass all available machine conditions, while all the subspaces do not overlap so that the ambiguity in determining the machine condition can be avoided.

Feature extraction is essentially an information redundancy reduction procedure which has also been argued to be the underlying principle of biological sensory systems [3]. Sparse coding was advocated as an effective mathematical

description for such a principle [4], which could possibly confer several advantages: it allows for increased storage capacity in associative memories; it makes the structure in natural signals explicit; it represents complex data in a way that is easier to read out at subsequent levels of processing; and it saves energy [5]. These advantages still hold when it is used as a signal processing technique. Sparse coding has been well established as a feature extraction technique and has been applied for typical pattern recognition tasks [6]. In this paper, sparse coding is taken for vibration analysis and feature extraction.

Most resulting machine condition spaces are unobservable, because the dimension of the resulting features is always still high, even after dimension reduction. Although a trained linear or nonlinear classifier may automate the determination of the machine condition, the visualization of the machine condition space with the monitored data mapped on would give equipment maintainers a transparent understanding about the current machine condition. Self-organizing map (SOM) [7] is a natural choice to achieve this goal. It is an unsupervised artificial neural network (ANN), which can capture the topology or probability distribution of input data and present them with map units. In this paper, the SOM is not only used for visualization of the machine condition space. At the same time, for an unknown machine condition, it can be automatically detected with a prespecified criterion such as the minimum quantization error (MQE). This is necessary in the practical implementations of most CBM systems, because it is always impossible to construct a machine condition space encompassing all the conditions of a machine in the working environment.

Traditional visual methods for inspecting the machine condition use the spectrum analysis methods, such as the FFT-based spectrum, holospectrum, time-frequency distributions, etc. A review of the most commonly used spectrum analysis methods can be found in [8]. All these methods raise a high requirement on the professional quality of maintainers. By cascading sparse coding (SC) and self-organizing map (SOM), this paper proposes a new model (abbr. SC-SOM) for robust visual monitoring of machine condition. The objective is to devise a direct and easy-to-understand display of the machine condition. A benchmark bearing data set is used for verification. The remainder of this paper is organized as follows: in Section 2, the preliminaries about sparse coding and self-organizing map are reviewed; in Section 3, the SC-SOM model is discussed; in Section 4, the experiment to verify the SC-SOM model is given; conclusions and discussions are given in Section 5.

## 2 Preliminaries

### 2.1 Sparse Coding

**The Sparse Coding Model.** The standard sparse coding model is based on a linear generative model. A signal  $\mathbf{x} = [x_1, x_2, \dots, x_M]^T$  can be described as:

$$\mathbf{x} = \mathbf{D}\mathbf{s} + \boldsymbol{\epsilon} = \sum_{k=1}^K \mathbf{d}_k s_k + \boldsymbol{\epsilon} \quad , \quad (1)$$

where  $\mathbf{D} \in \mathbb{R}^{M \times N}$  is a matrix called *dictionary*, of which the column,  $\mathbf{d}_k \in \mathbb{R}^M$ , is the basis function (a.k.s. *atom*);  $\mathbf{s} = [s_1, s_2, \dots, s_K]^T$  is the *sparse representation* of the input signal  $\mathbf{x}$ ; and  $\boldsymbol{\epsilon}$  is assumed to be Gaussian residual noise.

Given a dictionary  $\mathbf{D}$ , sparse coding is to determine the sparse representation of the input signal  $\mathbf{x}$ . Comparing with PCA and ICA, the dictionary for sparse coding is always overcomplete, i.e.  $M < K$ , and in most cases,  $M \ll K$ . When  $\mathbf{D}$  is a full-rank matrix, there could be infinite number of solutions available for  $\mathbf{s}$ . However, sparsity is desired in solving the coefficients  $s_k$ , i.e., most of the coefficients  $s_k$  are zero or nearly zero. A more formal and expressive definition of sparse coding is:

$$\min_{\mathbf{s}} \|\mathbf{s}\|_0 \quad \text{subject to} \quad \|\mathbf{x} - \mathbf{D}\mathbf{s}\|_2 \leq \gamma, \quad (2)$$

here the  $\ell^0$  norm,  $\|\cdot\|_0$ , is introduced as a sparsity measure, which counts the nonzero entries in a vector; the reconstruction accuracy is assessed by the  $\ell^2$  norm,  $\|\cdot\|_2$ ;  $\gamma$  is the approximation error tolerance.

**The Choice of the Dictionary.** Conventional signal processing techniques, such as FFT-based methods, wavelet-based methods, and so on, rely on a pre-specified dictionary for signal decomposition. This is simple and always leads to fast algorithms. However, these hand-selected basis functions are limited in their abilities in adapting to different types of data [9]. And, the success of such dictionaries in applications depends on how suitable they are to sparsely describe the signals in question [10].

A much more flexible approach is to learn the dictionary from the data themselves so that each basis function can capture a maximal amount of structures in data, i.e., dictionary learning. Formally, given a batch of (i.i.d) signals  $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ , dictionary learning is about to learn a dictionary  $\mathbf{D}$  adapting to the analyzed data with respect to the model in (2).

**Shift-Invariant Sparse Coding.** In the practical implementations of most signal processing techniques, one long signal sequence is always truncated into smaller blocks. The same feature in a long signal may be shifted at any location within a block due to the different truncating lengths. With the standard sparse coding model, the same feature shifted at different locations may be regarded as different basis functions during the dictionary learning process. The shift-invariant sparse coding model can avoid this by modeling the same feature at all possible locations:

$$\mathbf{x} = \sum_{k=1}^K \mathbf{d}_k * \mathbf{s}_k + \boldsymbol{\epsilon}, \quad (3)$$

where a convolution operator  $*$  is used to succinctly reuse each basis function  $\mathbf{d}_k$  at all possible shifts within the signal  $\mathbf{x}$ ,  $\mathbf{s}_k \in \mathbb{R}^{M-K+1}$  is the coefficient vector associated with basis function  $\mathbf{d}_k$  ( $\mathbf{s}_k$  is denoted as the *sparse activation* of  $\mathbf{d}_k$  in this paper).



For both sparse coding and shift-invariant sparse coding, exact determination of the sparse representation is a NP-hard problem [11]. Approximate algorithms are considered instead, such as matching pursuit (MP) [12], basis pursuit (BP) [13], etc. While for dictionary learning, various algorithms are also developed. In this paper, both the efficient dictionary learning and coefficients solving algorithms developed in [14] for shift-invariant sparse coding are adopted. Given a set of training signals  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ , assuming a uniform prior over the bases, a maximum a posterior (MAP) estimator was proposed for both the bases and coefficients in [14]:

$$\min_{\mathbf{d}, \mathbf{s}} \sum_{i=1}^N (\|\mathbf{x}_i - \sum_{k=1}^K \mathbf{d}_k * \mathbf{s}_{k,i}\|_2^2 + \beta \sum_{k=1}^K \|\mathbf{s}_{k,i}\|_1) , \tag{4}$$

$$\text{subject to} \quad \|\mathbf{d}_k\|_2^2 \leq c, 1 \leq k \leq K , \tag{5}$$

where  $\mathbf{s}_{k,i}$  is the sparse activation of  $\mathbf{d}_k$  for signal  $\mathbf{x}_i$ ;  $c$  is a constant constraint preventing  $\mathbf{d}_k$  from being very large while  $\mathbf{s}_{k,i}$  being very small;  $\beta$  is an optimization penalty parameter;  $\|\cdot\|_1$  is introduced here as a sparsity measure. The objection function (4) is although not jointly convex, but convex with one variable while holding the other fixed. Thus, two convex optimization problems are formulated for solving coefficients and bases respectively. For more details about these algorithms, please refer [14].

### 2.2 Self-Organizing Map

The SOM is an unsupervised neural network developed by Kohonen in [7]. A basic SOM model consists of an input layer and a mapping layer. Usually, the mapping layer with  $m$  map units, i.e., the neurons, is a 2-dimensional rectangular or hexagonal neuron grid with all adjacent neurons fully connected. During the training of a SOM, each input sample vector  $\mathbf{x}$  is mapped onto the mapping layer. The distances between the input sample vector  $\mathbf{x}$  and the weight vectors  $\mathbf{w}_i$  of all neurons are calculated using some distance measures (such as the Euclidian distance). The neuron with the weight vector closest to  $\mathbf{x}$  is chosen as the best matching unit (BMU) of  $\mathbf{x}$ . Then, its topological neighbors are updated so that they are moved closer to the input vector. The updating rule is as follows:

$$\mathbf{w}_i(t + 1) = \mathbf{w}_i(t) + \alpha(t)h(n_{\text{BMU}}, n_i, t)(\mathbf{x} - \mathbf{w}_i(t)) \quad (i = 1, \dots, m) , \tag{6}$$

where  $h(n_{\text{BMU}}, n_i, t)$  is the neighborhood function around the BMU, which is usually chosen as a Gaussian function;  $\alpha(t)$  is the learning rate.

The training of a SOM is a competitive learning process, also known as *vector quantization*. The distance between an input sample vector and the weights of its BMU is the minimum quantization error (MQE):

$$\text{MQE} = \|\mathbf{x} - \mathbf{w}_{\text{BMU}}\| . \tag{7}$$

After training with a batch of samples, the SOM forms a semantic map, in which the adjacent neurons share close weights. In this way, the similar samples

would be mapped close together and the dissimilar apart. According to the input samples of which the labels are known, the neurons on the semantic map can be labeled. The SOM provides some useful tools for visual inspection of the trained semantic map, which are very useful for machine condition monitoring. Among them, the most often used is the *unified distance matrix* (U-matrix), which shows the distances between adjacent neurons. Others include the component plane, hit histogram, trajectory, and so on.

### 3 The SC-SOM Model

By cascading sparse coding and the SOM, two manifold learning methods are in fact cascaded. Based on an overcomplete dictionary, sparse coding produces a more simple flattened representation of the curved manifold structure of the data [5]. In addition, sparse coding in this model decomposes signals based on basis functions learned from the signals themselves, consistent sparse representations are able to be solved when some disturbances interfere, benefiting from the good adaptability of the learned dictionary. So structured and consistent features are able to be formulated, which further contributes to the robust monitoring of machine condition. The SOM nonlinearly projects the resulting sparse features for machine condition space visualization and maintenance decision-making support. The implementation of the SC-SOM model for machine condition monitoring is presented in Fig. 1. As it is shown, two stages are needed: offline training and online monitoring.

During the offline training stage, an overcomplete dictionary is firstly constructed by merging several sub-dictionaries with each sub-dictionary learned from one class of historical signals; secondly, by solving the sparse representations of all the historical signals with the constructed dictionary, sparse features are formulated from them; finally, the MCM (a semantic map representing the machine condition space, abbr. MCM) is trained by self-organizing mapping all the sparse features. Then, the constructed dictionary and the trained MCM are

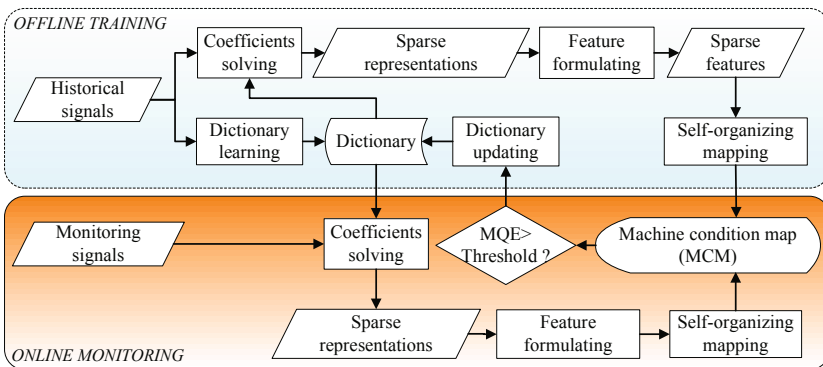
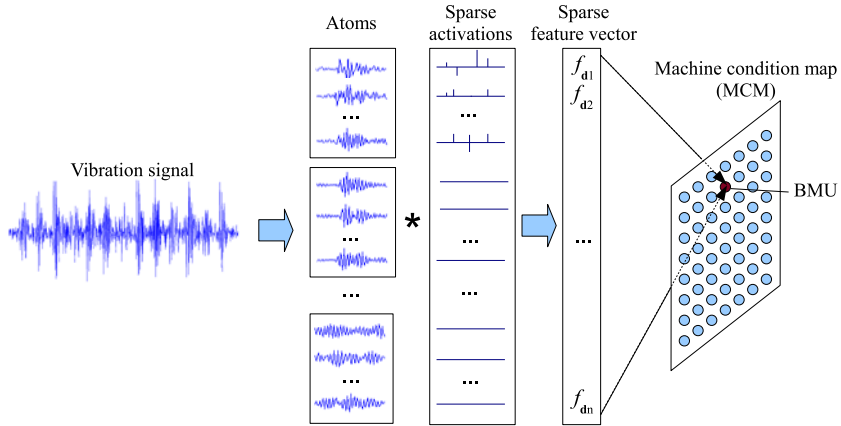


Fig. 1. The SC-SOM model



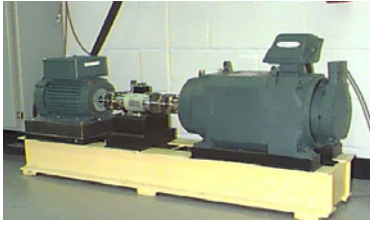
**Fig. 2.** Schematic description of the SC-SOM model with vibration signal

used for online monitoring of machine condition: solving the sparse representations with the offline constructed dictionary, formulating the sparse features as offline training did and self-organizing mapping the sparse features, the monitoring signals are finally mapped onto the MCM. Considering the diversity of signal types used in CBM, the generic sparse coding is described in the SC-SOM model. For the widely used signal form, time-series, the shift-invariant sparse coding is always employed. A schematic description of the SC-SOM model for online monitoring with vibration signals is presented in Fig. 2, in which shift-invariant sparse coding is used for vibration signal decomposition.

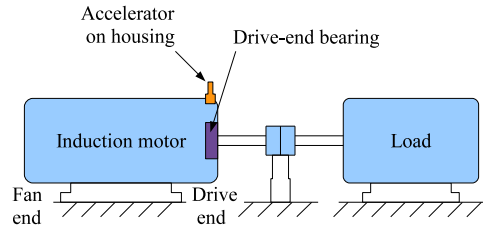
Generally speaking, the MCM can be read in such a way: if the label of the mapped BMU is known, the label tells the current machine condition; otherwise, checking the MQE of the BMU, if the value of the MQE is less than a prespecified threshold, the current machine condition can be referred from the labels of its adjacent units. The MQE value exceeding the prespecified threshold means a new machine condition rising. In this case, the dictionary needs to be updated by learning a new sub-dictionary from the current monitoring signals and merging it into the dictionary.

## 4 Robust Visual Monitoring of Rolling Element Bearings

Bearings are the most commonly used parts in mechanical systems, and vibration monitoring is the most popular approach to inspect the bearing condition. As a demonstrative application and verification of the SC-SOM model, the benchmark bearing vibration data from Case Western Reserve University [15] are chosen for evaluation.



(a) Picture of the bearing test stand by Loparo [15]



(b) Schematic diagram of the test stand

**Fig. 3.** The bearing test stand

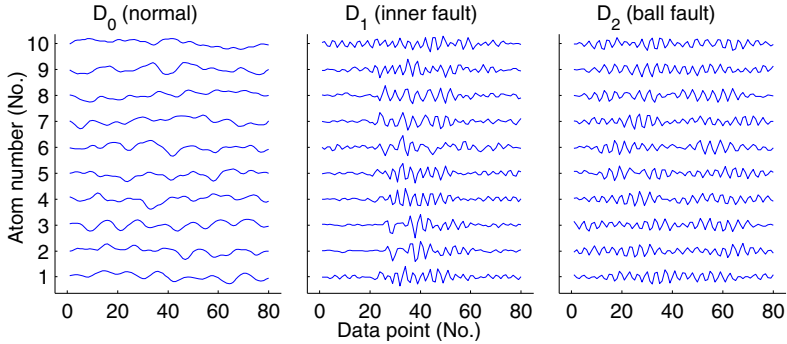
#### 4.1 Experiment Setup

The test stand of rolling element bearings is shown in Fig. 3. On the left was a 2 hp, three-phase induction motor (Reliance Electric 2HP IQPreAlert motor). A dynamometer (right) was connected to the motor through a torque sensor (center) by self-aligning coupling. The dynamometer was used to make sure that the desired torque load levels could be achieved. The bearings under test (SKF 6205-2RS JEM deep groove ball bearing) were the ones on the drive end side supporting the motor shaft. Single point faults were introduced to the bearings using electro-discharge machining with fault diameter of 7 mils. Fault locations were on inner race, ball, and outer race. Tests were carried out under four levels of motor loads (0, 1, 2 and 3 hp). An accelerator was mounted on the motor housing at the drive end of the motor. Vibration signals were collected using a 16-channel DAT recorder at a sampling frequency of 12 K Hz.

Totally 3787 samples of signal were analyzed, which included the signals sensed from both normal (abbr. N) and the three seeded fault conditions: inner race fault (abbr. IF), ball fault (abbr. BF), outer race fault (abbr. OF) under the four motor loads (0, 1, 2, 3 hp). Each sample signal had 1024 sample points. Following the SC-SOM model presented in Fig. 1, half of the sample signals (291 samples) under motor load 0 hp with three bearing conditions: normal, IF, BF, were chosen for dictionary learning and the MCM training. The rest of the sample signals were used to simulate the online monitoring signals. Among them, the sample signals of OF fault were chosen for the novel fault detection.

#### 4.2 Offline Training

**Dictionary Learning.** Among the 291 samples for offline training, there were 146 samples of normal condition and about 72 samples for the IF and BF conditions respectively. The shift-invariant dictionary learning algorithms developed in [14] were adopted. Atoms were learned from the three classes of signals separately. The length of one atom was specified with 80 to capture one impulse under the fault conditions, and the number of atoms was specified with 10. Taking into consideration that each atom may appear at any location within a



**Fig. 4.** The learned atoms from normal, inner fault and ball fault conditions

1024-point signal block, the 10 atoms essentially compose a 10x overcomplete dictionary ( $10 \times 1024$  atoms).

The learned atoms are shown in Fig. 4. Examining the learned atoms, we can notice that: apparent impulses appear in the atoms of IF condition; continuous small impulses are the main pattern for the BF condition; while for the normal condition, the atoms are much smoother. This is consistent with the mechanisms of generating defective impulses in bearings that: under the IF condition, the rolling elements strike the crack on the inner raceway periodically; while under the BF condition, the cracked rolling element strikes the inner and outer raceways continuously with random angles by the seeded hole.

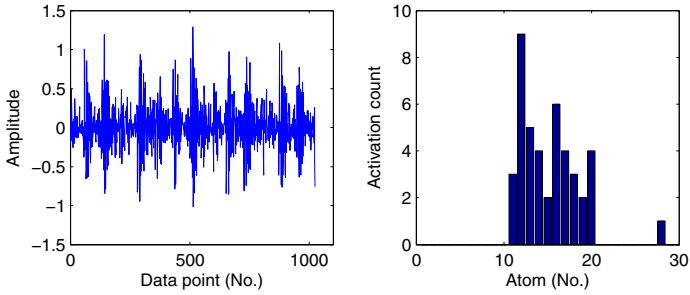
By merging these three dictionaries, a redundancy dictionary is formed:

$$\mathcal{D} = [\mathbf{D}_0, \mathbf{D}_1, \mathbf{D}_2] = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_{30}] , \quad (8)$$

where  $\mathbf{d}_1$  is the first atom in  $\mathbf{D}_0$ ,  $\mathbf{d}_2$  is the second atom in  $\mathbf{D}_0$ , and so on.

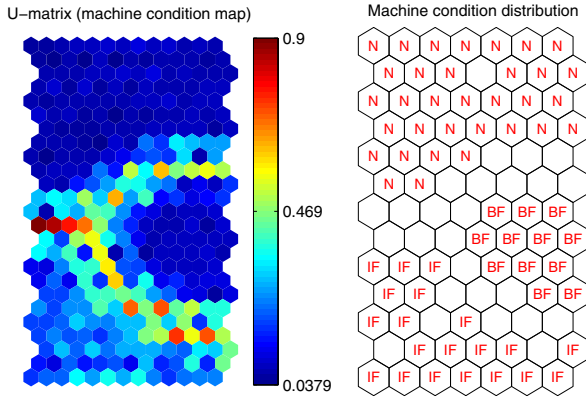
**Feature Extraction.** Given the dictionary  $\mathcal{D}$ , the sparse representations of all the sample signals can be solved. Note that each sub-dictionary  $\mathbf{D}_i$  is learned from one class of bearing condition. For one class of sample signals, the atoms learned from them are prone to be activated to approximate the original signals. One example is given in Fig. 5. The sample signal shown on the left is chosen from the IF condition under motor load 0 hp. The number of activations of each atom in its sparse representation is counted and shown on the right. We can see that the atoms in the sub-dictionary  $\mathbf{D}_1$  are mainly activated in the signal's sparse representation.

Thus, a 30-dimensional feature vector can be formulated from the sparse representation of each sample signal, with each entry in the feature vector calculated from the sparse activation of each atom. The calculation of the feature can be varied, such as the number of activations, the absolute values, the sum of squares, and so on. In our evaluations, the sum of squares was chosen.



**Fig. 5.** The number of activations of the atoms (right) in approximating the signal (left)

**Machine Condition Map (MCM) Training.** Training a SOM neural network with all the extracted sparse features from the training samples, a MCM (as shown on the left of Fig. 6) was generated. The MCM is a typical U-matrix, which represents the distances between neurons. Three subspaces can be easily recognized from the MCM. According to the bearing conditions represented by the training signals, the neurons can also be labeled. The labeled result is shown on the right of Fig. 6.



**Fig. 6.** A MCM trained with normal (N), inner race fault (IF) and ball fault (BF) condition

It is important to notice that, the label information should be taken as a reference to judge the machine condition, but not the only. The adjacent neurons with similar colors on the MCM share close distances so that they can be interpreted with the same machine condition. In the practical working environments, deviations of the mapped BMU on the MCM may occur due to various disturbances. The maintainer should hold a tolerance in reading the MCM.

### 4.3 Online Monitoring

**Machine Condition Recognition.** After the offline training, the constructed dictionary  $\mathcal{D}$  and the trained MCM are to be used for online monitoring.

Firstly, the test signals under motor load 0 hp (the other half except for training, totally 293 samples) were tested. The sparse features were extracted following the routine as described in subsection 4.2. All the extracted sparse features from normal, IF and BF conditions were mapped on the MCM. The hit histograms are plotted in Fig. 7a, in which bigger hexagon means more signals are mapped on it. It can be read from the hit histograms that the test signals are mapped to their corresponding subspaces. Although some of the mapped neurons are unlabeled, their meanings can be referred from their adjacent neurons.

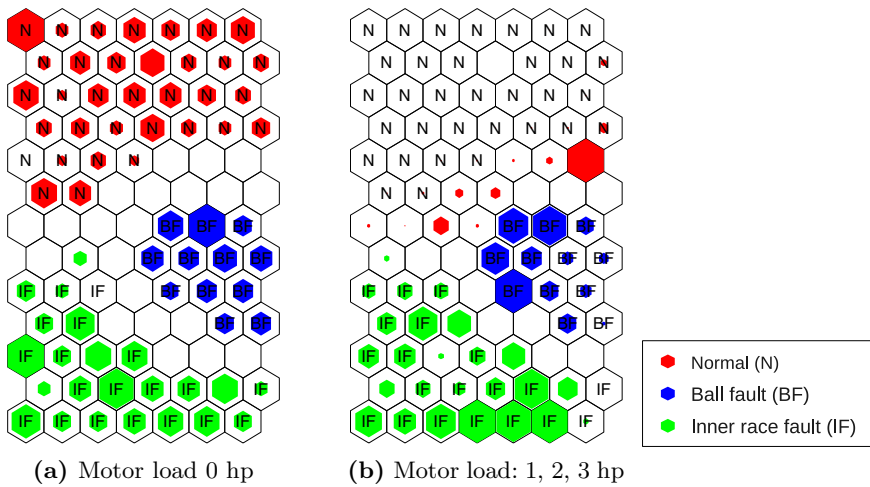
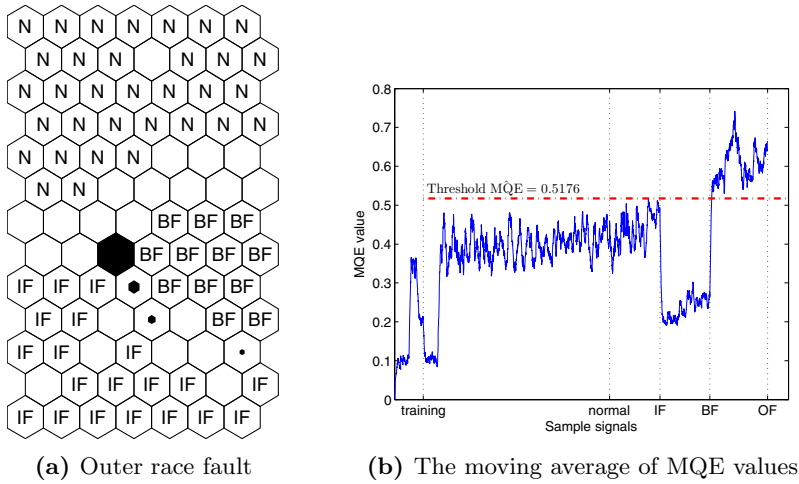


Fig. 7. Machine condition recognition with the trained MCM

Secondly, the sample signals from other motor loads (motor load 1, 2, 3 hp, totally 2619 samples) were tested. The varying motor loads can be regarded as the disturbances. Sparse features were extracted and mapped on the MCM with the same way. The hit histograms are plotted in Fig 7b. We can see that the mappings of IF and BF condition signals still distribute to their subspace. The mappings of the normal condition signals deviate slightly from the labeled normal subspace but still distribute around it. So, the bearing conditions can also be apparently judged.

**Novel Fault Condition Detection.** For a novel fault condition detection test, the sample signals from OF condition with the four different motor loads were used. Totally 584 samples were mapped on the trained MCM (see Fig. 6). The mapping results are plotted in Fig. 8a with hit histograms. As it is shown, the mappings leave the normal subspace and deviate from the IF and BF subspaces.



**Fig. 8.** A novel fault condition (outer race fault) detection example

The minimum quantization error (MQE) can be referred for validation of a novel fault condition. The moving average of MQE for all the 3787 samples is plotted in Fig. 8. On the plot, the MQE of the training samples are firstly plotted, followed by the test samples of four bearing conditions. The plot shows that the MQE goes highest with the sample signals of OF condition. With the prespecified MQE threshold  $MQE = 0.5176$ , the emergence of a novel fault can be determined. In the practical implementations, the threshold value of MQE can be firstly set as the largest MQE in the training samples and further revised if a monitoring signal is mapped on a labeled neuron.

### 5 Conclusions and Discussions

By cascading sparse coding and the SOM, a model named SC-SOM is proposed for robust visual monitoring of machine condition. Sparse coding is used for feature extraction, while the SOM is used for machine condition space visualization and novel fault detection. Two stages are needed in implementing the model. After offline training, a dictionary and a machine condition map (MCM) is formed for online monitoring. The application on bearing condition monitoring shows that the proposed SC-SOM model is an effective condition monitoring tool. More important is that it is robust to disturbances and possesses novel fault detection ability.

The proposed SC-SOM model can be implemented as an ecological condition monitoring system. The learned dictionary can be interpreted as a machine condition dictionary. Dictionary learning is the core of the SC-SOM model. In this paper, only the robust visual monitoring ability of the SC-SOM model is evaluated. The updating of the dictionary deserves a special research.



**Acknowledgments.** This research is supported by the National Science and Technology Major Project in China (Grant no. 2009ZX04014-103) and the National High Technology Research and Development Program of China (863 Program, Grant no. 2008AA042801, and 2009AA043000).

## References

1. He, Q., Kong, F., Yan, R.: Subspace-based gearbox condition monitoring by kernel principal component analysis. *Mechanical Systems and Signal Processing* 21, 1755–1772 (2007)
2. Sun, Q., Chen, P., Zhang, D., Xi, F.: Pattern recognition for automatic machinery fault diagnosis. *Journal of Vibration and Acoustics* 126, 307–316 (2004)
3. Barlow, H.: Possible principles underlying the transformations of sensory messages. *Sensory Communication*, 217–234 (1961)
4. Földiák, P.: Forming sparse representations by local anti-hebbian learning. *Biological Cybernetics* 64, 165–170 (1990)
5. Olshausen, B.A., Field, D.J.: Sparse coding of sensory inputs. *Current Opinion in Neurobiology* 14, 481–487 (2004)
6. Wright, J., Yang, A., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 210–227 (2009)
7. Kohonen, T.: *Self-Organizing Maps*, 3rd edn. Springer, Heidelberg (2001)
8. Jardine, A.K.S.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing* 20, 1483–1510 (2006)
9. Lewicki, M.S., Sejnowski, T.J.: Learning overcomplete representations. *Neural Computing* 12, 337–365 (2000)
10. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing* 54, 4311–4322 (2006)
11. Davis, G., Mallat, S., Avellaneda, M.: Adaptive greedy approximations. *J. Construct. Approx.* 13, 57–98 (1997)
12. Mallat, S.G., Zhang, Z.: Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing* 41, 3397–3415 (1993)
13. Chen, S., Dohono, D., Saunders, M.: Atomic decomposition by basis pursuit. *SIAM Review* 43, 129–159 (2001)
14. Grosse, R., Raina, R., Kwong, H., Ng, A.Y.: Shift-invariant sparse coding for audio classification. In: *UAI* (2007)
15. Loparo, K.: Bearing vibration data set. Case Western Reserve University (2003), <http://www.eecs.cwru.edu/laboratory/bearing/download.htm>

# An Evolving Machinery Fault Diagnosis Approach Based on Affinity Propagation Algorithm

Nan Li, Yanming Li, Haining Liu, and Chengliang Liu

School of Mechanical Engineering, Shanghai Jiaotong University, Shanghai 200240, China  
{linansjtu, ymli, hnliu, chl1iu}@sjtu.edu.cn

**Abstract.** An evolving approach combining unsupervised clustering and supervised classification for intelligent machinery fault diagnosis is proposed. As the key point of the approach, the unsupervised clustering module for detecting a novel fault is specified in this paper. Incorporating with prior information in the historical data, a constrained clustering method is developed based on the affinity propagation (AP) algorithm. The clustering method is validated through experimental case studies on the bearing fault diagnosis. The results show that the clustering method has the self-learning abilities to detect the novel faults and with the evolving abilities of the proposed approach, one can start with just the normal condition data and continue building the diagnosis scheme as the new fault events occur.

**Keywords:** fault diagnosis, fault detection, unsupervised clustering, supervised classification, affinity propagation.

## 1 Introduction

Machinery fault diagnosis is a procedure of mapping the features obtained in the feature space to machine faults in the fault space. This mapping process is essentially a problem of pattern recognition. Traditionally, fault recognition is done manually with auxiliary graphical tools such as power spectrum graph, phase spectrum graph, wavelet phase graph, etc. However, costly training and highly skilled personnel are needed because manual recognition requires expertise in the specific area of the diagnostic application [1]. Therefore, automatic fault recognition is highly desirable. This can be achieved by intelligent data analysis based on the features extracted from the signals.

In the field of data analysis two terms frequently encountered are supervised and unsupervised methods of data classification and clustering methodologies. While supervised methods mostly deal with training classifiers for known symptoms, unsupervised clustering provides exploratory techniques for finding hidden patterns in the data.

The research of intelligent fault diagnosis methods is primarily dominated by supervised algorithms, like neural networks, decision trees, support vector machines, etc [2-4]. These supervised fault classification methods work accurately and efficiently for trained signatures only and generate erroneous results when encounter novel data not belonging in the domain of training data. This leads to lower acceptance of these methods especially in the industry, where obtaining labeled data for different operating conditions may be hard.

While supervised classification refers to learning through labeled examples, unsupervised clustering refers to analyzing and extracting inductive patterns from the data without a trained model. Owing to the superiority in dealing with uncertainty and independence from supervisors, unsupervised clustering algorithms, such as k-means, fuzzy c-means, SOM, etc [5, 6], have been widely studied and applied to fault diagnosis. One limitation of the existing clustering algorithms is that there is not much predefined information for clustering in most cases. Some unsupervised methods require a predefined number of clusters before they perform clustering, but it is not always possible to know the number in advance. Therefore, the algorithms frequently result in inexact clustering results. Another problem is that these algorithms may emerge problems of time complexity when dealing with large scale data. This is not suitable for the online machinery fault detection due to the large amount of data in the industry.

Also, it is important to recognize that the archived data does not possibly encompass all possible fault scenarios. Ideally one would like to build a diagnosis scheme based on existing historical data and update the knowledge database as new faults arise during online machinery fault detection. Therefore, the fault diagnosis system also needs to have evolving ability, i.e., when deployed online it should be able to identify the occurrence of new faults and establish relevant signatures or patterns representing the novel fault.

In this paper, we propose an evolving approach which combines unsupervised clustering and supervised classification for intelligent fault diagnosis. In order to guarantee classification accuracy and improve robustness for novel faults detection, before the supervised classification module, an unsupervised clustering module is introduced into the whole fault diagnosis loop. Thus, the unsupervised module gives the provision for dealing with novel faults in the system while the supervised module can be updated when a novel fault is presented. Fig. 1 shows the schematic diagram of the proposed approach.

Once the data is processed in real time, it is regarded as the present data. Then the unsupervised clustering is operated on the combined data sets of the present data and the historical data. If the number of clusters obtained through the unsupervised clustering matches with the known classes based on the historical data, then the existing classifying system is used. However when there is mismatch and the number of clusters obtained is greater than the number of trained classes, a temporary label for the present data is created. This temporary label can then be efficiently labeled by expert personnel. After it is confirmed as the novel fault finally, the supervised classification and the unsupervised clustering will be both retrained and updated for the incremented data sets. Having detected the novel fault, the new cluster information can be merged with the existing knowledge database and used for future fault diagnosis. In principle, with the evolving abilities of the proposed approach, one can start with just the normal condition data and continue building the diagnosis scheme as the new fault events occur.

The supervised classification module for accurate fault classification is trained from historical data. There are a variety of classification methods developed in machine learning and pattern recognition community such as neural networks, decision trees, support vector machines, etc. The supervised classifier which has been studied in a large amount of available literature will not be discussed in this paper.

The unsupervised clustering module for detecting a novel fault is the key point of the whole diagnosis approach and will be discussed in detail in this paper.

The rest part of this paper is organized as follows. In Section 2, the unsupervised clustering module based on the affinity propagation (AP) algorithm is briefly introduced. Section 3 presents the results with discussions of the application to bearing fault diagnosis. Section 4 draws the conclusion.

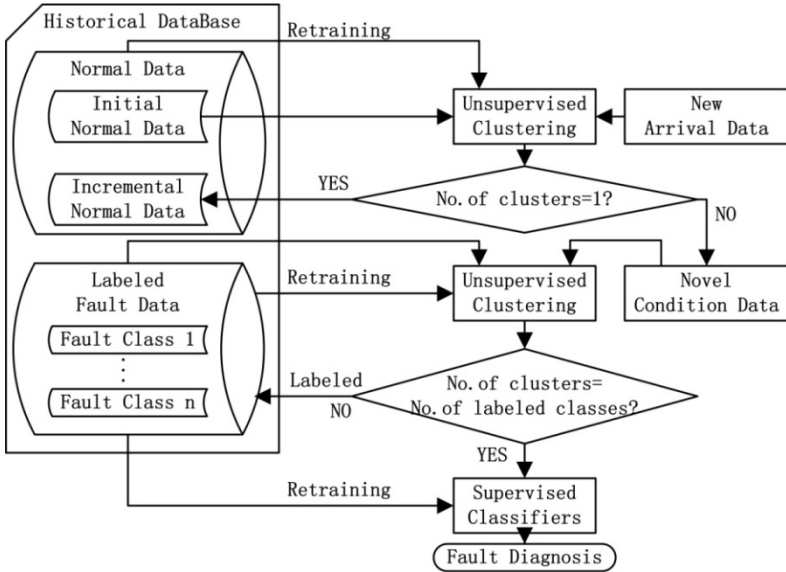


Fig. 1. Schematic diagram of the proposed approach

## 2 The Unsupervised Clustering Module Based on AP Algorithm

The algorithm adopted in the unsupervised clustering module is based on affinity propagation (AP) algorithm which was published in Science, 2007 [7]. AP is an innovative and readily-extensible clustering algorithm that identifies exemplars quickly and successfully. Reported in the original paper, AP achieves a significant improvement compared with those conventional clustering methods like K-means, spectral clustering and super-paramagnetic clustering. One advantage of AP is that the number of exemplars need not be specified beforehand. It consistently achieves comparable or better results in far less time for large datasets. Because of its splendid performance, AP has been successfully applied to many fields, such as image clustering, gene-expression data, text mining and large scale data. Drawn by the success of this algorithm from different fields we attempt to use it in the industry as a tool for machinery fault diagnosis. In order to fit the algorithm to the actual machinery fault diagnosis and obtain appropriate clustering results, we develop a constrained AP algorithm using prior information in the historical data.

### 2.1 AP Algorithm

The classical techniques for clustering, such as k-means clustering and fuzzy c-means clustering, are both very sensitive to the initial selection of cluster centers and usually must be re-run multiple times with different initializations in an attempt to find a good solution [8]. However, this works well only when the number of clusters is small and chances are good that at least one random initialization is close to a good solution.

Unlike the classical clustering algorithms, AP simultaneously considers all data points as potential exemplars, and exchanges real-valued messages between data points so as to find a subset of exemplar points that best describe the data.

AP takes input as a so called similarity matrix **S**, where its element  $s(i, j)$  indicates that how well the data point  $i$  is suited to be the exemplars for data point  $j$ . The similarity should be set according to the optimization criteria of the problem, and actually, it needn't to be symmetrical.

There are two kinds of message exchanged between data points, “responsibility”  $r(i, k)$ , sent from point  $i$  to candidate exemplar  $k$ , reflects the accumulated evidence for how well-suited point  $k$  is to serve as the exemplar for point  $i$ , taking into account other potential exemplars for point  $i$ . The “availability”  $a(i, k)$ , sent from candidate exemplar point  $k$  to point  $i$ , reflects the accumulated evidence for how appropriate it would be for point  $i$  to choose point  $k$  as its exemplar, taking into account the support from other points that point  $k$  should be an exemplar.

The availabilities are initialized to be zero, and then the responsibilities and availabilities are computed by coupling according to the updating rules as follows [7]:

$$r(i, k) = s(i, k) - \max_{k' \neq k} (a(i, k') + s(i, k')) \tag{1}$$

$$a(i, k) = \min\{0, r(k, k) + \sum_{i' \in \{i, k\}} \max(0, r(i', k))\} \tag{2}$$

$$a(k, k) = \sum_{i' \neq k} \max(0, r(i', k)) \tag{3}$$

At any point during AP, availabilities and responsibilities can be combined to identify exemplars. The message-passing procedure may be terminated after a fixed number of iterations, after changes in the messages fall below a threshold, or after the local decisions stay constant for some number of iterations. AP requires only simple local computations that are easily implemented and is capable of achieving better results with far less CPU time than most of conventional methods.

### 2.2 The Unsupervised Clustering Module

Based on passing messages between data points, AP searches for clusters through an iterative process until a high-quality set of exemplars and corresponding clusters emerges. Although the number of exemplars need not be specified beforehand, the appropriate number of exemplars depends on the input exemplar preferences parameter  $p$  which is in diagonal of similarity matrix **S**. The preference parameter  $p$  (its initial value is negative) indicates the preference that data point  $i$  be chosen as a

cluster center, and influences the output clusters and the number of clusters. However, there is no exact corresponding relation between the parameter  $p$  and the number of output clusters. Therefore, how to find an appropriate clustering solution is an unsolved problem when using AP algorithm.

In order to obtain appropriate clustering results in the unsupervised clustering module, we develop a constrained AP algorithm using prior information in the historical data. As mentioned in the proposed diagnosis approach, firstly, the AP algorithm is operated on the historical data sets in order to find the appropriate parameter  $p$  through the training methods, and then, the AP algorithm is applied with this parameter  $p$  to present added data for fault detection. If the novel fault is confirmed, the parameter  $p$  will be updated through the retraining from the current data sets.

The algorithm for finding the appropriate parameter  $p$  is executed by adjusting the value of  $p$  constantly with a certain step until the final clustering result can be consistent with the prior constrained information in the historical data sets, such as pairwise constraints of must-link or cannot-link points.

### 3 Application to Bearing Fault Diagnosis

#### 3.1 Case Study Description

In order to validate the proposed fault diagnosis approach, experimental analysis on rolling element bearings is conducted. The database used in the experimental study is taken from the Case Western Reserve University (CWRU) Bearing Vibration Data Center [9]. At this center, the experimental data were collected from ball bearings in the drive and fan ends of an induction motor that drives the mechanical system.

The test rig, shown in Fig.2, consists of a 2 hp, three-phase induction motor, a torque sensor and a dynamometer connected by a self-aligning coupling. The dynamometer is controlled so that desired torque load levels can be achieved. An accelerometer is mounted on the motor housing at the drive and fan ends. Motor bearings are faulted using the electro-discharge machining (EDM) method. The data collection system consists of a high bandwidth amplifier particularly designed for the vibration signals and a data recorder with a sampling frequency of 12,000 Hz per channel. Data are collected for four different conditions: normal condition, ball fault, inner race fault and outer race fault. For these cases, the faulted bearings are

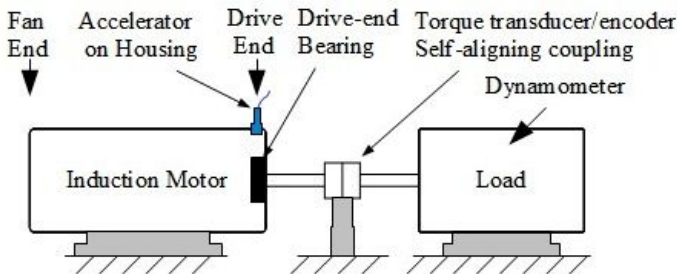


Fig. 2. Schematic diagram of the test rig

reinstalled and vibration data are collected for three different fault diameters (0.007, 0.014, and 0.021 inches). All of the experiments are repeated for four different load conditions (0, 1, 2, and 3 HP). Therefore, the experimental data consist of four vibration signals for the normal condition and 12 vibration signals for each of the inner race, outer race, and ball fault conditions. The vibration signals belonging to each class are used to construct feature vectors.

### 3.2 Feature Extraction and Selection

In any fault diagnosis problem, feature extraction from the raw signals is a very critical step. Because these extracted features can not only characterize the information relevant to the bearing conditions, but also affects the final diagnosis results.

The faults that typically occur in roller bearings are usually caused by local defects which generate a series of impact vibrations and cause the vibration amplitude to increase. It is shown that the peak value and root mean square reflect the vibration amplitude and energy, and the kurtosis value, crest factor, shape factor and impulse factor characterize the impact existing in the roller bearings. The kurtosis value and crest factor are robust to varying operating conditions of the bearings, and are good indicators of incipient faults. The impulse factors are also good indicators of the sharp impulses generated by the contact of a defect with the bearing mating surfaces [10]. Therefore, we select these seven features, including mean, peak value, root mean square, kurtosis value, shape factor, impulse factor and crest factor, from each sample to represent the bearing conditions.

### 3.3 Validation of Novel Fault Detection through Unsupervised Clustering

For each time of fault detection, 25 samples are collected at a time and every sample contains 2048 data points. As the approximate motor speed is 1797 rpm, the motor rotates about 5 revolutions over the time interval of 2048 data points when the sampling frequency is 12,000Hz. The feature vectors with seven variables are constructed from the data points for each sample. In order to evaluate the proposed fault detection approach, 225 data samples covering four different operating conditions of normal condition, ball fault, inner race fault and outer race fault are selected from the datasets. The normal condition and each fault condition with the defect sizes of 0.007 inches are divided into training and testing instances. The detailed description of the data set used in this paper is shown in Table 1.

**Table 1.** Data set used in this paper

Number of training samples	Number of testing samples	Operation condition
50	25	normal
25	25	ball fault
25	25	inner race fault
25	25	outer race fault

### 3.3.1 Step 1: Novel Condition Detection According to Normal Condition

As mentioned in the proposed approach, the first step is to build an unsupervised clustering module for detecting a novel condition since there is only normal condition in historical data initially.

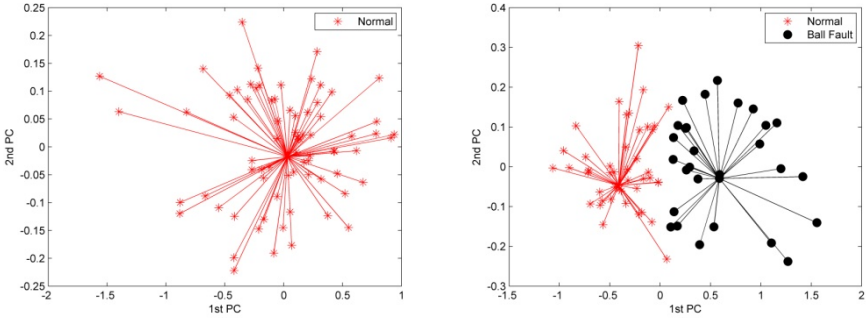


Fig. 3. Clustering of 50 training and 25 testing samples of normal condition (left); 50 training samples of normal condition and 25 testing samples of ball fault (right)

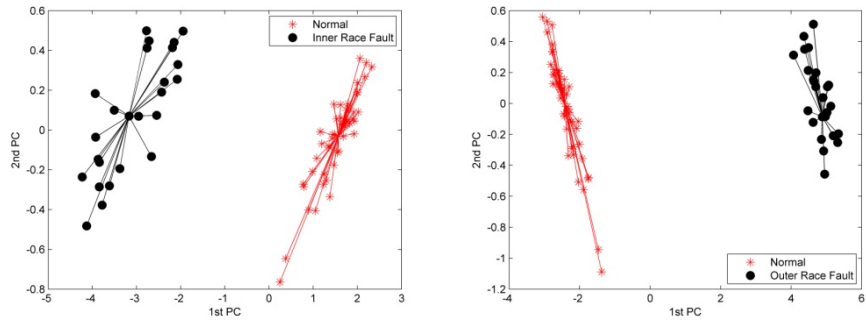


Fig. 4. 50 training samples of normal condition and 25 testing samples of inner race fault (left); 50 training samples of normal condition and 25 testing samples of outer race fault (right)

First of all, 50 training samples of the historical normal condition are used to find the appropriate  $p$  in AP algorithm for the clustering. Then 25 testing samples of the normal, ball fault, inner race fault or outer race fault are respectively added to 50 training samples of historical normal condition in order to validate the availability of the parameter  $p$  for the novel condition detection.

For visualization of the feature vectors with seven dimensions, we implement the principal component analysis (PCA) method on the clustering results obtained by the proposed algorithm. The plot of the first two principal components (PCs) of the results is shown in the following figures.

In the left of Fig.3, there is only one cluster during the clustering of 50 training and 25 testing samples of normal condition. While in the right of Fig.3 and in Fig.4, there



are two clusters during the respective clustering of 50 training samples of normal condition and three types of fault data (ball fault, inner race fault and outer race fault) with 25 testing samples in each type. The preference parameter  $p$  is set to -16.9107 and the CPU times taken to carry out this algorithm are 3.624, 2.412, 2.581 and 2.774s, respectively.

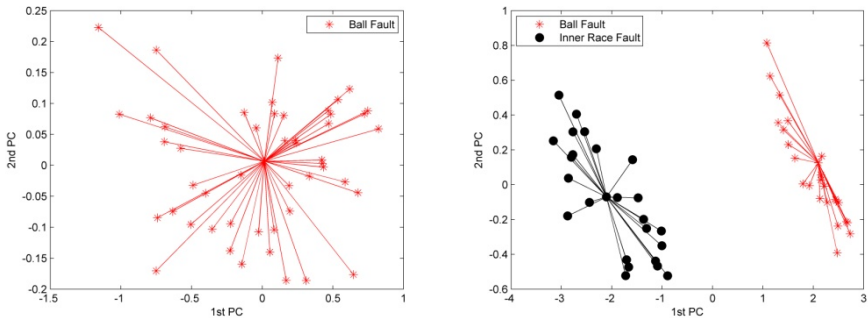
After this step, the normal condition data will be added to the incremental normal dataset which can be retrained for the update of the clustering parameter  $p$ , and the novel condition data will be processed in the next unsupervised clustering step.

**3.3.2 Step 2: New Fault Type Detection According to Labeled Fault Type**

The clustering method is operated on the combination of the novel condition data from the first step and the fault data labeled in the historical fault dataset. If the number of clusters obtained through the clustering method matches with the known classes of the historical fault data, then the existing classifying system is used. However when there is mismatch and the number of clusters obtained is greater than the number of trained classes, a temporary label for the present data is created. After it is confirmed as the novel fault finally, the existing classifying system and the unsupervised clustering will be both retrained and updated for the incremented fault dataset. In order to validate this method, we assume that the ball fault with 25 samples has already been obtained in the historical dataset.

First of all, 25 training samples of ball fall are used to find the appropriate  $p$  for the clustering. Then 25 testing samples of ball fault and inner race fault are respectively added to 25 training samples of ball fault in order to validate the availability of the parameter  $p$  for the new fault type detection.

As shown in the left of Fig.5, there is still one cluster during the clustering of 25 training and 25 testing samples of ball fault while in the right of Fig.5, there are two clusters during the clustering of 25 samples of ball fault and 25 testing samples of inner race fault. This indicates that the new fault type is detected. The preference parameter  $p$  is set to -17.6148 and the CPU times taken to carry out this algorithm are 1.886 and 1.653s, respectively.

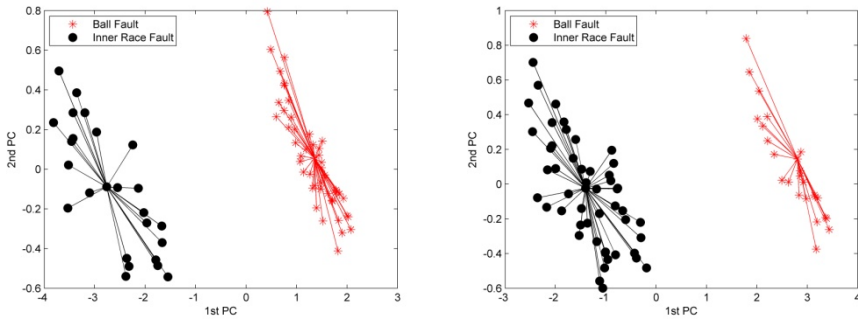


**Fig. 5.** 25 training samples and 25 testing samples of ball fault (left); 25 training samples of ball fault and 25 testing samples of inner race fault (right)

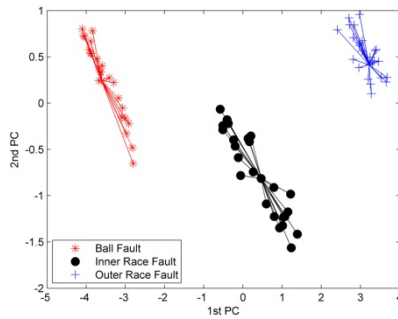
### 3.3.3 The Evolving Ability

When the novel fault is detected and the historical fault dataset is updated to 50 samples within 25 samples of ball fault and 25 samples of inner race fault. Then the parameter  $p$  is adjusted for the new condition. After the update, if there are 25 testing samples of ball fault or inner race fault obtained again, the number of the clusters is still two as shown in Fig.6. This means that these two types of fault have already existed in the historical fault dataset and the accurate classification will be operated by the supervised classifier trained from the historical dataset. The preference parameter  $p$  is set to -20.2674 and the CPU times taken to carry out this algorithm are 3.366 and 3.743s, respectively.

In order to testify the ability of the novel fault detection of the updated clustering model, 25 testing samples of outer race fault are added to the fault dataset within 50 samples of ball fault and inner race fault. As shown in Fig.7, there are three clusters during the clustering which indicates that another novel fault is detected. This means that the proposed unsupervised clustering method can evolve continuously by self-learning for the novel fault detection. When novel fault data not belonging in the



**Fig. 6.** Clustering of 25 training samples of ball fault, 25 training samples of inner race fault and 25 testing samples of ball fault (left); 25 training samples of ball fault, 25 training samples of inner race fault and 25 testing samples of inner race fault (right)



**Fig. 7.** Clustering of 25 training samples of ball fault, 25 training samples of inner race fault and 25 testing samples of outer race fault

domain of existing training datasets arrive, this unsupervised method will find these hidden patterns and update for the new condition while other traditional supervised fault classification methods, such as SVM, may generate erroneous results. With the evolving abilities of the proposed approach, one can start with just the normal condition data and continue building the diagnosis scheme as the new fault events occur.

## 4 Conclusions

This paper presents an evolving approach combining unsupervised clustering and supervised classification for intelligent machinery fault diagnosis. A constrained affinity propagation (AP) algorithm is developed for the unsupervised clustering module which is the critical component in the whole approach. The clustering algorithm is validated through experimental case studies on the bearing fault diagnosis. By adjusting the preference parameter through learning from prior information in historical data, a diagnosis scheme is built gradually. The testing results show that the proposed method has the self-learning ability to detect the novel faults and with the evolving abilities it can be easily implemented to the online faults detection.

## Acknowledgments

This work is supported by National Science and Technology Major Project in China (Grant no. 2009ZX04014-103) and the National High Technology Research and Development Program of China (863 Program, Grant no. 2008AA042801, and 2008AA042803).

## References

1. Jardine, A.K.S., Lin, D., Banjevic, D.: A Review on Machinery Diagnostics and Prognostics Implementing Condition-based Maintenance. *Mechanical Systems and Signal Processing* 20, 1483–1510 (2006)
2. Siddique, A., Yadava, G.S., Singh, B.: Applications of Artificial Intelligence Techniques for Induction Machine Stator Fault Diagnostics: Review. In: *Proceedings of the IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, pp. 29–34. IEEE Press, New York (2003)
3. Lo, C.H., Wong, Y.K., Rad, A.B.: Intelligent System for Process Supervision and Fault Diagnosis in Dynamic Physical Systems. *IEEE Transactions on Industrial Electronics* 50, 581–592 (2006)
4. Widodo, A., Yang, B.S.: Support Vector Machine in Machine Condition Monitoring and Fault Diagnosis. *Mechanical Systems and Signal Processing* 21, 2560–2574 (2007)
5. Wong, M.L.D., Jack, L.B., Nandi, A.K.: Modified Self-organizing Map for Automated Novelty Detection Applied to Vibration Signal Monitoring. *Mechanical Systems and Signal Processing* 20, 593–610 (2006)
6. Li, R., Chen, J., Wu, X., Alugongo, A.A.: Fault Diagnosis of Rotating Machinery Based on SVD, FCM, and RST. *Int. J. Adv. Manuf. Technol.* 27, 128–135 (2005)

7. Frey, B.J., Dueck, D.: Clustering by Passing Messages between Data Points. *Science* 315, 972–976 (2007)
8. Liao, T.W.: Clustering of Time Series Data: a survey. *Pattern Recognition* 38, 1857–1874 (2005)
9. Case Western Reserve University Bearings Vibration Data Center,  
<http://www.eecs.cwru.edu/laboratory/bearing/download.htm>
10. Howard, I.M.: A Review of Rolling Element Bearing Vibration Detection, Diagnosis and Prognosis. Defense Science and Technology Organization, Australia (1994)

# Closed Form Solution for the Scale Ambiguity Problem in Monocular Visual Odometry

Isaac Esteban, Leo Dorst, and Judith Dijk

University of Amsterdam and TNO, The Netherlands

**Abstract.** This paper presents a fast monocular visual odometry algorithm. We propose a closed form solution for the computation of the unknown scale ratio between two consecutive image pairs. Our method requires only 1 2D-3D correspondence. A least square solution can also be found in closed form when more correspondences are available. Additionally we provide a first order analysis on the propagation of the error from the noise in the image features to the computation of the scale. We show by means of simulated and real data that our method is more robust and accurate than standard techniques. We demonstrate that our visual odometry algorithm is well suited for the task of 3D reconstruction in urban areas.

## 1 Introduction

Accurate estimation of the ego-motion of a vehicle is an essential step for autonomous operation. This estimation can be performed using a wide range of sensors such as GPS, wheel encoders, laser scanners or cameras. The use of visual information is in particular interesting due to the flexibility of the cameras, their reduced cost and the possibility of integration with complementary systems such as road lane warnings, pedestrian detectors or obstacle avoidance techniques.

Early work on motion estimation is based on stereo vision systems [14]. In this setup two cameras are located parallel to each other while recording the scene. Given the knowledge of the baseline and relative orientation of the cameras, one can obtain accurate results when estimating the ego-motion of the camera [5] [17]. More recent work is focussed on monocular vision systems. The problem of estimating the motion of a camera and reconstructing the 3D structure is called "structure from motion" [3] [10]. Successful results have been obtained using both omnidirectional and perspective cameras [20] [19] [17] [18].

Related to the ego-motion estimation is the process of Simultaneous Localization and Mapping, or SLAM [6]. It aims at estimating both the motion of the moving vehicle and the surrounding map. SLAM methods usually involve the use of several sensors, though recent developments have been made that use only cameras [8]. In this work we develop a robust motion estimation algorithm for monocular cameras that can be integrated with complementary mapping techniques such as SLAM.

The nature of monocular systems induces a scale ambiguity in the estimation of the motion and the reconstructed 3D scene (see Figure 1). The global scale cannot be recovered unless information about the real world is introduced. This is achieved, for instance, by gathering information from a GPS or about the distance travelled by the camera. Additionally, if the motion is estimated on a frame-to-frame basis, considering only the image features, there is a scale ambiguity between the estimated translation vectors. We call this the local scale problem. There are a number of solutions in the literature to solve the scale implicitly. Scaramuzza et al. [20] use information about the camera’s height to the ground and position with respect to the axis of the vehicle to obtain the global scale factor. Also, in [19] a direct/average observation of the speed of the vehicle is used to account for the distance travelled by the camera. These techniques, used extensively in the literature, suffer from a number of disadvantages. When introducing the speed of the camera, or the distance travelled, we are forced to use external devices such as a GPS or a speedometer. If the speed is averaged, then the motion of the camera is limited to a steady rate.

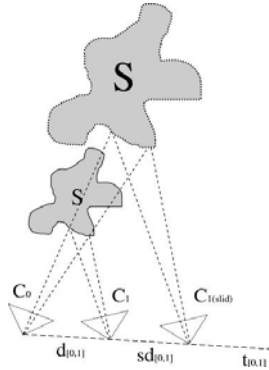
An alternative common technique, used in [17] [18], is to solve the camera pose using 2D-3D correspondences instead of image only 2D-2D correspondences. This is the so called PnP (Pose from n Points) problem, for which a number of solutions exist [4] [1] [15] [2]. When using PnP methods to estimate the camera pose, the scale is implicitly estimated. These methods however suffer from severe error propagation. Due to the fact that the full camera motion is estimated using only 2D-3D correspondences, the error is propagated as follows: first, from the image features in cameras  $C_i$  and  $C_{i+1}$  to the estimation of camera pose at  $C_{i+1}$ . Second, the error is also propagated to the triangulation of the 3D points. Finally, the error is propagated to the estimation of the camera pose at  $C_{i+2}$ . This effect is propagated to subsequent camera poses and accumulates over time.

In this paper we focus on the local scale problem for which we derive an explicit closed form solution. We propose an alternative technique with the same computational complexity as the linear P6P solution, or related methods, but with considerable improvement in accuracy. Additionally, we provide a first order analysis on the propagation of the error and present a set of experiments with both simulated and real outdoors data.

This paper is organized as follows. In section 2 we present the notation and definitions. Section 3 presents our closed form solution for the computation of the local scale. In Section 4 we provide a first order analysis on the propagation of the error from the image noise and camera poses to the computation of the scale. In Section 5 we present our results on both simulated and real data. Finally, in Section 6 we draw some conclusions and point out future work.

## 2 Definitions and Notation

We employ the pinhole camera model, though the techniques we present can be accommodated to single view omnidirectional cameras. We begin with the definition of the estimated image features, 3D points and camera poses as a



**Fig. 1.** Scale problem in monocular vision. Estimating the camera pose  $C_1$  is intrinsically scale free. By sliding the camera over an unknown amount along the translation direction  $t_{[0,1]}$ , both structures  $S$  become indistinguishable reconstructions, only the scale is different.

disturbance of the true values. These definitions are useful to understand the derivation of the closed form solution for the local scale and later for the analysis on the error propagation. Note that at this point we do not model the error and only reason about how the error is propagated.

### 2.1 Disturbed Image Features

We define the disturbed images features  $\mathbf{x}_\alpha = \bar{\mathbf{x}}_\alpha + \Delta\mathbf{x}_\alpha$  in homogeneous coordinates (denoted by  $H$ ) as:

$$\mathbf{x}_{\alpha H} + \Delta\mathbf{x}_{\alpha H} = \begin{bmatrix} \mathbf{x}_\alpha \\ 1 \end{bmatrix} + \begin{bmatrix} \Delta\mathbf{x}_\alpha \\ 0 \end{bmatrix} = \begin{bmatrix} u_\alpha + \Delta u_\alpha \\ v_\alpha + \Delta v_\alpha \\ 1 \end{bmatrix}, \tag{1}$$

where  $\mathbf{x}_\alpha$  is the estimated location of the true image feature  $\bar{\mathbf{x}}_\alpha$  and  $\Delta\mathbf{x}_\alpha$  is the disturbance.

### 2.2 Disturbed Point in Space

Our method for the computation of the scale requires at least 1 2D-3D correspondence. We define a disturbed point in 3D space  $\mathbf{X}_\alpha = \bar{\mathbf{X}}_\alpha + \Delta\mathbf{X}_\alpha$ , or in homogeneous coordinates:

$$\mathbf{X}_{\alpha H} + \Delta\mathbf{X}_{\alpha H} = \begin{bmatrix} \mathbf{X}_\alpha \\ 1 \end{bmatrix} + \begin{bmatrix} \Delta\mathbf{X}_\alpha \\ 0 \end{bmatrix} = \begin{bmatrix} a_{x_\alpha} + \Delta a_{x_\alpha} \\ a_{y_\alpha} + \Delta a_{y_\alpha} \\ a_{z_\alpha} + \Delta a_{z_\alpha} \\ 1 \end{bmatrix} \tag{2}$$

Note that the disturbance represented here will depend on the error in the motion estimation, the feature location and the method for triangulation.

### 2.3 Disturbed Translation

We represent the translation direction with a unitary vector  $\mathbf{t}$  with the constraint  $\|\mathbf{t}\| = 1$ , or equivalently  $\mathbf{t}^T \mathbf{t} = 1$ . We impose this constraint in the disturbed translation  $\mathbf{t} = \bar{\mathbf{t}} + \Delta\mathbf{t}$ :

$$1 = (\bar{\mathbf{t}} + \Delta\mathbf{t})^T (\bar{\mathbf{t}} + \Delta\mathbf{t}) \tag{3}$$

$$1 = 1 + \bar{\mathbf{t}}^T \Delta\mathbf{t} + \Delta\mathbf{t}^T \bar{\mathbf{t}} + \Delta\mathbf{t}^T \Delta\mathbf{t} \tag{4}$$

Neglecting the second order terms ( $\Delta\mathbf{t}^T \Delta\mathbf{t} \approx 0$ ) we obtain that  $\Delta\mathbf{t}^T \bar{\mathbf{t}} = 0$ . This indicates that the disturbance  $\Delta\mathbf{t}$  is perpendicular to the unitary translation direction  $\bar{\mathbf{t}}$ . Expressed in coordinates:

$$\bar{\mathbf{t}}_\alpha = \begin{bmatrix} t_x + \Delta t_x \\ t_y + \Delta t_y \\ t_z + \Delta t_z \end{bmatrix}, \tag{5}$$

with  $\Delta t_x^T t_x + \Delta t_y^T t_y + \Delta t_z^T t_z = 0$ .

### 2.4 Disturbed Rotation

Given that most algorithms use a rotation representation based on a  $3 \times 3$  matrix  $R$ , we represent the error as a relative disturbance over the true rotation [12]:

$$R = \bar{R}(I + \Delta R) \tag{6}$$

A rotation matrix  $R$  is an orthogonal matrix ( $R^{-1} = R^T$ , or equivalently  $I = R^T R$ ). The disturbed rotation must be also an orthogonal matrix, therefore we can obtain, to first order:

$$I = (\bar{R}(I + \Delta R))^T (\bar{R}(I + \Delta R)) \tag{7}$$

$$I = \bar{R}^T \bar{R} + \bar{R}^T \bar{R} \Delta R + \Delta R^T \bar{R}^T \bar{R} + \Delta R^T \bar{R}^T \bar{R} \Delta R \tag{8}$$

$$I \approx I + \Delta R + \Delta R^T \tag{9}$$

$$\Delta R^T = -\Delta R \tag{10}$$

Then  $\Delta R$  is a skew symmetric matrix of the form:

$$\Delta R = \begin{bmatrix} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \\ \omega_2 & -\omega_1 & 0 \end{bmatrix}, \tag{11}$$

where  $\omega$  is an axis-angle [12] vector representation for a small rotation (the disturbance of the rotation). For shortness we define the rotation matrix with row vectors:

$$R = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \tag{12}$$



The absolute disturbance  $\bar{R}\Delta R$  is also represented with with row vectors  $\Delta \mathbf{r}_1^T$ ,  $\Delta \mathbf{r}_2^T$  and  $\Delta \mathbf{r}_3^T$ .

So far we have described our notation with a first order analysis on the disturbances over image features, points in space, translation directions and rotations. Based on these definitions, we now describe our scale consistent motion estimation and then analyze the propagation of the error.

### 3 Robust and Scale Consistent Motion Estimation

We propose an algorithm for robust and scale consistent estimation of the motion of a camera through space. We present the algorithm in two parts. The first part deals with robust frame-to-frame motion estimation. This estimation is scale free. The second part deals with the estimation of the scale, which we solve in closed form.

#### 3.1 Robust Frame-to-Frame Motion

The first part of the algorithm deals with robustly estimating the frame-to-frame motion. This is a well known problem and there are a large number of solutions. Among the most used is the 5-point algorithm [16] and normalized 8-point algorithm [10]. In our work we choose to use the latter given its simplicity, though our contribution on the computation of the scale can be combined with any frame-to-frame motion estimation algorithms. Additionally, we employ a local iterative refinement step in order to improve the estimate without a large computational burden.

We summarize the steps of the algorithm:

- Extract SIFT [13] features using VLfeat [22].
- Match features using nearest neighbors in the SIFT descriptor space.
- Reject outliers using RANSAC [9].
- Estimate the essential matrix defining the frame-to-frame motion using the normalized 8-point algorithm [10].
- Obtain a rotation  $R$  and unitary translation  $t$  using the method by Horn [11].
- Refine the estimate by minimizing the reprojection error using Levenberg-Marquardt. In this setup the internal calibration parameters (focal length, pixel ratio and focal point) are also optimized.

In general no more than 4 iterations of the refinement step are needed for convergence. The optimization of the rotation is performed in the normalized quaternion space. Note that this optimization step can cope with some of the shortcomings of the 8-point algorithm such as degeneracies. Given a good set of inliers and a slightly larger number of iterations, the procedure can converge to a good solution even when all points lie in a plane. This motion estimation procedure yields a rotation matrix  $R$  and a unitary translation direction  $t$ .

---

<sup>1</sup> Given that the true rotation is unknown, the absolute disturbance can be approximated by multiplying the estimated rotation  $R$  with the relative disturbance  $\Delta R$ .

### 3.2 Computation of the Local Scale

Using the above algorithm we can robustly estimate the motion of the camera through space. However, the translation ratios between different image pairs is still unknown. In this section we derive a closed form solution for the computation of the local scale.

We know [10], given the camera model, that the relation between the image features and the reconstructed 3D points in homogeneous coordinates (denoted by  $H$ ) is defined by:

$$m\mathbf{x}_{\alpha H} = P X_{\alpha H}, \quad (13)$$

where  $P$  is the scale free camera motion ( $P = [R|\mathbf{t}]$  with  $\|\mathbf{t}\| = 1$ ) and  $m$  is the scaling necessary to obtain image points and not image plane points. Given that the camera motion, obtained as described in section 3.1, is only calculated up to scale, we introduce the scaling factor  $s$ :

$$m\mathbf{x}_{\alpha H} = [R|s\mathbf{t}]X_{\alpha H}, \quad (14)$$

or equivalently:

$$\begin{bmatrix} mu_{\alpha} \\ mv_{\alpha} \\ m \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & st_x \\ R_{21} & R_{22} & R_{23} & st_y \\ R_{31} & R_{32} & R_{33} & st_z \end{bmatrix} \begin{bmatrix} a_{x\alpha} \\ a_{y\alpha} \\ a_{z\alpha} \\ 1 \end{bmatrix} \quad (15)$$

For every 2D-3D feature match  $\alpha$ , we obtain three equations:

$$mu_{\alpha} = \mathbf{r}_1^T \mathbf{X}_{\alpha} + st_x \quad (16)$$

$$mv_{\alpha} = \mathbf{r}_2^T \mathbf{X}_{\alpha} + st_y \quad (17)$$

$$m = \mathbf{r}_3^T \mathbf{X}_{\alpha} + st_z \quad (18)$$

**Note:** At this point we can solve this system for  $s$  in different ways. However, let us continue our derivation and we refer the reader to Section 3.3 for a detailed motivation.

In order to solve for  $s$ , we substitute equation 18 in equation 16 to remove the image scale factor  $m$ . Manipulating the terms we single out the scale parameter  $s$ :

$$(\mathbf{r}_3^T \mathbf{X}_{\alpha} + st_z)u_{\alpha} = \mathbf{r}_1^T \mathbf{X}_{\alpha} + st_x \quad (19)$$

$$(t_z u_{\alpha} - t_x)s = (\mathbf{r}_1^T - \mathbf{r}_3^T u_{\alpha})\mathbf{X}_{\alpha} \quad (20)$$

This is precisely of the form  $\mathbf{A}s = \mathbf{b}$ , where:

$$\mathbf{A} = [t_z u_{\alpha} - t_x] \quad (21)$$

$$\mathbf{b} = [(\mathbf{r}_1^T - \mathbf{r}_3^T u_{\alpha})\mathbf{X}_{\alpha}], \quad (22)$$

We can construct the vectors  $\mathbf{A}$  and  $\mathbf{b}$  using one row for every 2D-3D correspondence  $\alpha$ , and solve for  $s$  using SVD (Singular Value Decomposition), obtaining a solution in the least square sense. Given that  $\mathbf{A}$  and  $\mathbf{b}$  are vectors, we can also obtain a closed form solution. We want to obtain the scale  $s$  in the least

square sense, which means minimizing  $\|\mathbf{A}s - \mathbf{b}\|^2$ . We do so obtaining the partial derivative w.r.t.  $s$  and setting it to zero:

$$\|\mathbf{A}s - \mathbf{b}\|^2 = \mathbf{A}^T \mathbf{A}s^2 - 2\mathbf{A}^T \mathbf{b}s + \mathbf{b}^T \mathbf{b} \tag{23}$$

$$\frac{\partial(\|\mathbf{A}s - \mathbf{b}\|^2)}{\partial s} = 2\mathbf{A}^T \mathbf{A}s - 2\mathbf{A}^T \mathbf{b} = 0 \tag{24}$$

$$s = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} \tag{25}$$

Hence we obtain a linear solution in the least square sense for the scale parameter  $s$ . In order to simplify, we express the solution of equation 25 as:

$$s = \mathbf{A}^+ \mathbf{b} \tag{26}$$

where:

$$\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \tag{27}$$

### 3.3 A Note on Alternative Methods

Given the set of 3 equations in 16, 17 and 18, there are 4 different ways to solve the system for  $s$ .

- **Method 1:** solve using equations 16 and 18 as we do. This implies using  $u_\alpha$  and  $t_x$ .
- **Method 2:** alternatively, the system can be solved using equations 17 and 18. This implies using  $v_\alpha$  and  $t_y$ .
- **Method 3:** combining both, the system can be solved using methods 1 and 2, stacking the solutions into vectors  $\mathbf{A}$  and  $\mathbf{b}$  (resulting in vectors that are twice as long).
- **Method 4:** The last possibility is using equations 16 and 17, combining  $u_\alpha$ ,  $v_\alpha$ ,  $t_x$  and  $t_y$  in a single solution. This solution is similar to methods 1 and 2 but using the ratio  $\frac{u_\alpha}{v_\alpha}$  (or the inverse).

It is at first sight difficult to chose one particular method. In an urban environment where the camera moves along a street, most of the motion occurs in the  $x$  direction, while the motion on the  $y$  axis is mostly negligible. In this case, we can motivate our decision by suggesting that the motion in the direction  $y$  will not contribute to the solution  $s$ . Rather it will disturb it since the motion will be the result of noise. Naturally, if the camera moves mainly in the  $y$  direction, the alternative method 2 should be employed. However our derivation and subsequent analysis holds by simply substituting  $t_x$  by  $t_y$  and  $\mathbf{r}_1^T$  by  $\mathbf{r}_2^T$  (along with the corresponding errors). Furthermore, our analysis also holds (or can be easily adjusted) for methods 3 and 4. For method 3, the resulting error will be result of stacking errors computed for method 1 and 2. For method 4, the error on the ratio  $\frac{u_\alpha}{v_\alpha}$  needs to be used instead of  $\Delta u_\alpha$ . In order to strengthen our motivation, we simulate a camera moving in an urban environment. In our setup the motion of the camera occurs mainly in the  $x$  direction. We implement all 4 methods and compare the accuracy in the estimation of the scale. Figure 3 shows the results.

## 4 Error Propagation in Scale Computation

Now that we have a closed form solution to compute the local scale, we wish to analyze how the error is propagated. We begin with the first order error propagation in the computation of  $s$ . To obtain a full picture of the propagation, the subsequent error analysis is performed on the computation of  $\mathbf{A}^+$ ,  $\mathbf{A}$  and  $\mathbf{b}$  independently. To first order, the error in equation 26 is:

$$\Delta s = \mathbf{A}^+ \Delta \mathbf{b} + \Delta \mathbf{A}^+ \mathbf{b} \quad (28)$$

This is however expressed in terms of  $\Delta \mathbf{A}^+$  and  $\Delta \mathbf{b}$ . We need to obtain those expressions as a function of the original disturbances in image features, points in space and camera motion.

### 4.1 Error Propagation in the Computation of $\mathbf{A}^+$

We obtain an expression for the computation of  $\mathbf{A}^+$  in equation 27. We now analyze how the error is propagated in its computation. We begin by introducing the disturbances:

$$\mathbf{A}^+ + \Delta \mathbf{A}^+ = ((\mathbf{A} + \Delta \mathbf{A})^T (\mathbf{A} + \Delta \mathbf{A}))^{-1} (\mathbf{A} + \Delta \mathbf{A})^T \quad (29)$$

$$= (\mathbf{A}^T \mathbf{A} + \mathbf{A}^T \Delta \mathbf{A} + \Delta \mathbf{A}^T \mathbf{A} + \Delta \mathbf{A}^T \Delta \mathbf{A})^{-1} (\mathbf{A} + \Delta \mathbf{A})^T \quad (30)$$

By first order Taylor, we may derive a first order approximation for the disturbance in the computation of  $\mathbf{A}^+$ :

$$\Delta \mathbf{A}^+ = -(\mathbf{A}^T \mathbf{A})^{-1} ((\mathbf{A}^T \Delta \mathbf{A} + \Delta \mathbf{A}^T \mathbf{A})(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T + \Delta \mathbf{A}^T) \quad (31)$$

#### Proof

Neglecting some second order terms in equation 30, we obtain:

$$\mathbf{A}^+ + \Delta \mathbf{A}^+ \approx (\mathbf{A}^T \mathbf{A} + \mathbf{A}^T \Delta \mathbf{A} + \Delta \mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A} + \Delta \mathbf{A})^T \quad (32)$$

We define:

$$Z = (\mathbf{A}^T \mathbf{A} + \mathbf{A}^T \Delta \mathbf{A} + \Delta \mathbf{A}^T \mathbf{A})^{-1} \quad (33)$$

Now, we need to transform  $Z$  in the form  $Z = (\mathbf{A}^T \mathbf{A})^{-1} + C$  so we can obtain a description of the disturbance  $\Delta \mathbf{A}^+$ . To accomplish this we employ a first order approximation using Taylor series. We first define:

$$a = \mathbf{A}^T \mathbf{A} \quad (34)$$

$$n = \mathbf{A}^T \Delta \mathbf{A} + \Delta \mathbf{A}^T \mathbf{A}. \quad (35)$$

We use these in equation 33 and transform it to obtain:

$$Z = (a + n)^{-1} = (a(1 + a^{-1}n))^{-1} = (1 + a^{-1}n)^{-1} a^{-1}. \quad (36)$$

We need to transform  $Z$  according to:

$$(a + n)^{-1} = a^{-1} + C \quad (37)$$

The first part  $((1 + a^{-1}n)^{-1})$  is approximated using a first order taylor series:

$$Z \approx (1 - a^{-1}n)a^{-1} = a^{-1} - a^{-1}na^{-1} = a^{-1} + (-a^{-1}na^{-1}), \tag{38}$$

where we obtain  $C = -a^{-1}na^{-1}$ . This provides a first order approximation of the error propagation in the computation of  $\mathbf{A}^+$ :

$$\begin{aligned} \mathbf{A}^+ + \Delta\mathbf{A}^+ &= ((\mathbf{A}^T \mathbf{A})^{-1} + C)(\mathbf{A} + \Delta\mathbf{A})^T \\ &= ((\mathbf{A}^T \mathbf{A})^{-1} - (\mathbf{A}^T \mathbf{A})^{-1}n(\mathbf{A}^T \mathbf{A})^{-1})(\mathbf{A} + \Delta\mathbf{A})^T \end{aligned} \tag{39}$$

$$\approx (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T - (\mathbf{A}^T \mathbf{A})^{-1} (n(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T + \Delta\mathbf{A}^T) \tag{40}$$

$$= \mathbf{A}^+ - (\mathbf{A}^T \mathbf{A})^{-1} ((\mathbf{A}^T \Delta\mathbf{A} + \Delta\mathbf{A}^T \mathbf{A})(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T + \Delta\mathbf{A}^T) \tag{41}$$

Therefore, the disturbance to first order that is propagated through the computation of  $\mathbf{A}^+$  is:

$$\Delta\mathbf{A}^+ = -(\mathbf{A}^T \mathbf{A})^{-1} ((\mathbf{A}^T \Delta\mathbf{A} + \Delta\mathbf{A}^T \mathbf{A})(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T + \Delta\mathbf{A}^T) \tag{42}$$

### 4.2 Error Propagation in Computation of $\mathbf{A}$

In the previous section we derive an expression for  $\Delta\mathbf{A}^+$  in terms of  $\Delta\mathbf{A}$ , now we obtain a first order expression for  $\Delta\mathbf{A}$ :

$$\mathbf{A} + \Delta\mathbf{A} = (t_z + \Delta t_z)(u_\alpha + \Delta u_\alpha) - (t_x + \Delta t_x) \tag{43}$$

$$\Delta\mathbf{A} = t_z \Delta u_\alpha + \Delta t_z u_\alpha - \Delta t_x \tag{44}$$

### 4.3 Error Propagation in Computation of $\mathbf{b}$

Finally, we need to obtain an expression for the error propagation in the computation of  $\mathbf{b}$ . We follow the same procedure as before, introducing the disturbances in equation 22 and propagating to first order:

$$\mathbf{b} + \Delta\mathbf{b} = [(\mathbf{r}_1^T + \Delta\mathbf{r}_1^T - (\mathbf{r}_3^T + \Delta\mathbf{r}_3^T)(u_\alpha + \Delta u_\alpha)) (\mathbf{X}_\alpha + \Delta\mathbf{X}_\alpha)] \tag{45}$$

$$\approx [(\mathbf{r}_1^T + \Delta\mathbf{r}_1^T - \mathbf{r}_3^T u_\alpha - \mathbf{r}_3^T \Delta u_\alpha - \Delta\mathbf{r}_3^T u_\alpha) (\mathbf{X}_\alpha + \Delta\mathbf{X}_\alpha)] \tag{46}$$

$$\approx [(\mathbf{r}_1^T - \mathbf{r}_3^T u_\alpha) \mathbf{X}_\alpha + (\mathbf{r}_1^T - \mathbf{r}_3^T u_\alpha) \Delta\mathbf{X}_\alpha + (\Delta\mathbf{r}_1^T - \mathbf{r}_3^T \Delta u_\alpha - \Delta\mathbf{r}_3^T u_\alpha) \mathbf{X}_\alpha] \tag{47}$$

$$\Delta\mathbf{b} = [(\mathbf{r}_1^T - \mathbf{r}_3^T u_\alpha) \Delta\mathbf{X}_\alpha + (\Delta\mathbf{r}_1^T - \mathbf{r}_3^T \Delta u_\alpha - \Delta\mathbf{r}_3^T u_\alpha) \mathbf{X}_\alpha] \tag{48}$$

### 4.4 Conclusions

In Section 4, in the same spirit as 12, we present a first order analysis on the error propagation when computing the local scale (see Section 3.2). We obtain a closed form expression for the propagated error (equation 28). In order to make use of the solution, we performed the analysis on the error propagation on the computation of  $\mathbf{A}^+$  (equation 42),  $\mathbf{A}$  (equation 44) and  $\mathbf{b}$  (equation 48). Getting all four equations together yields a first order closed form representation of the propagation of the error in the computation of the scale using any number of

2D-3D correspondences. This result can be used, for instance, to estimate the confidence on the computation of the scale. This, in turn, can be introduced in a SLAM (or any probabilistic mapping) approach where the covariance of the estimate can now be computed without the need of computationally expensive Monte Carlo simulations. Given that we provide an expression for error in the scale, the error in the 3D triangulation and the frame-to-frame motion needs to be provided. This however can be estimated using either Monte Carlo simulations or, given that the motion is scale independent, can be learned based on noise in the image features. Alternatively, the error propagation on the 8-point algorithm [21] can be employed together with our derivation.

## 5 Experiments and Results

### 5.1 Comparison with P6P

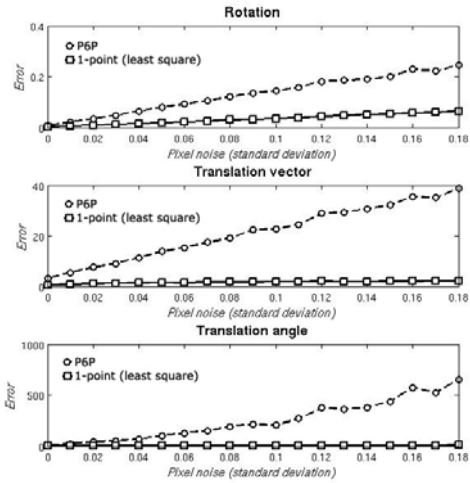
In this section we present our experimental results with a fully simulated scene of which 3 synthetic images are recorded. 30 3D points are projected onto the 3 cameras located at different positions. The internal parameters of the cameras are considered perfect. Gaussian noise is introduced only in the image feature location. We perform 2000 runs in which a different 3D scene is created and the cameras are randomly positioned. For each run, we estimate the motion of camera 3 using P6P [2] and our novel approach. We increment the noise in the image feature locations from 0.005 to 0.18 pixels of standard deviation. We then compare the estimated camera poses (rotation and translation) with the ground truth camera poses. The results are shown in figure 2. The plots show that our method outperforms the P6P reference algorithm by a large factor. For these experiments no global optimization is used. The improvement is particularly significant in the estimation of the translation, where our approach behaves more linearly. In this case, the P6P algorithm becomes almost unusable as the noise in the feature location increases. In order to reach a comparable accuracy, 4 times more 2D-3D correspondences are necessary. This cannot be solved by just increasing the frame rate as the number of images required to obtain the same accuracy becomes too large and so does the computational costs.

### 5.2 Outdoors Visual Odometry

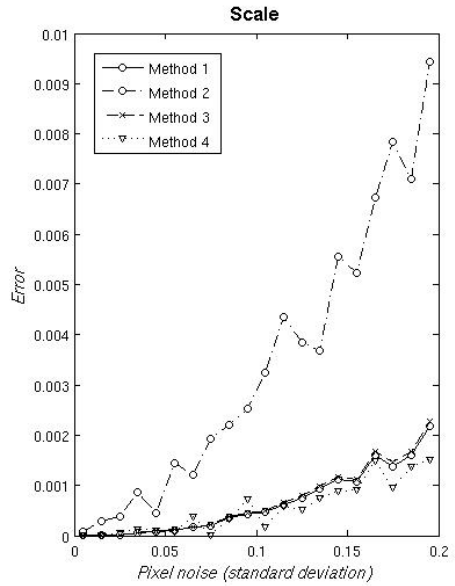
We also perform two experiments on real outdoors data. In both experiments images are recorded with a hand held digital reflex camera (DSLR). We use a wide angle lens at a resolution of 1728x1152 pixels. For the first set, 60 images (see figure 4) are recorded covering a total distance of 100 meters of an urban area, following a straight trajectory along one street. Results are shown in figure 5. For this experiment no drift can be perceived. This is because of our accurate

---

<sup>2</sup> The P6P reference algorithm is a linear algorithm based on the Direct Linear Transform method commonly used for comparison. We choose this algorithm given that, as our proposed method, it is linear and solves in a least square sense.



**Fig. 2.** LEFT: Error on the estimation of rotation and translation of a third camera pose given 2 previous camera poses and the reconstructed structure using linear triangulation (2000 runs). Top-left: error in the rotation calculated in the quaternion space. Middle-left: error in the translation vector calculated as the distance of the difference vector between ground truth and estimation. Bottom-left: error in the angle estimation of the translation direction.



**Fig. 3.** RIGHT: Error on the estimation of the scale using the 4 proposed methods (2000 runs). Method 2 is the worst given that the motion on the  $y$  direction does not contribute to the computation of  $s$ . Methods 1 and 3 offer similar performance, with a slight preference for method 1. Method 4 is slightly better than the rest. Our decision to use method 1 instead of 4 is motivated by the fact that method 4 can suffer from division by zero and its behavior is more erratic with a lower number of runs.

method and the fact that no Bundle Adjustment is used, which is a well known source for drift. The accurate results on the estimation of the camera motion are used to obtain an accurate reconstruction using [7]. For the second set, a total 174 images (see figure 6) are recorded covering a distance of 200+ meters. In this case the camera moves around a full block of houses. The same image is used for first and last position to ensure that the true last camera pose is exactly the same as where the first image is recorded. The trajectory is also recovered accurately without the aid of Bundle Adjustment. Given that the loop is closed, we can also measure the accuracy. The distance between the estimated position of the last image and the first image is below 1 meter. This represents an error of less than 0.5%.



Fig. 4. Sample images of first dataset

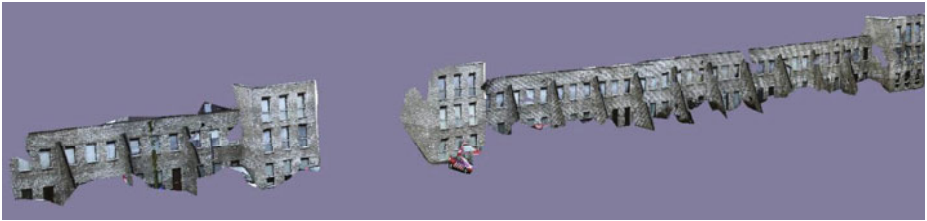
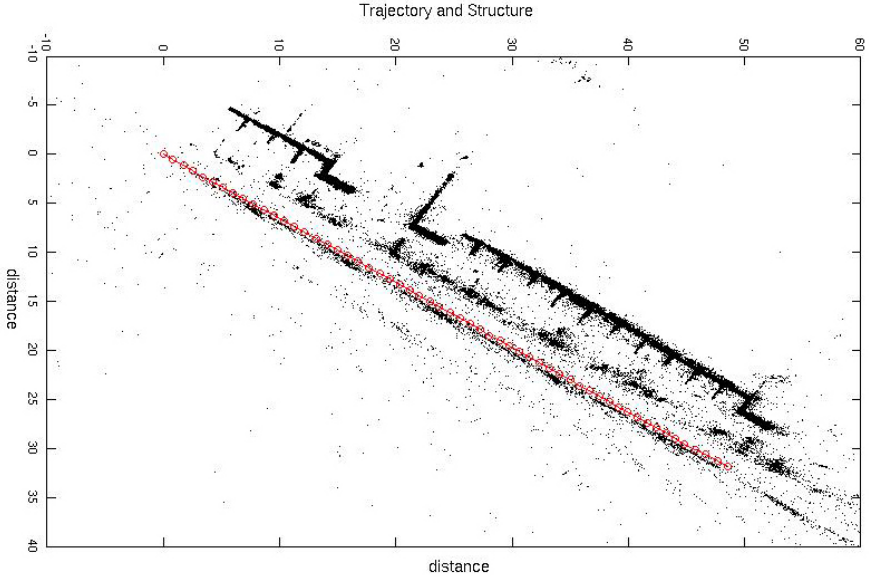
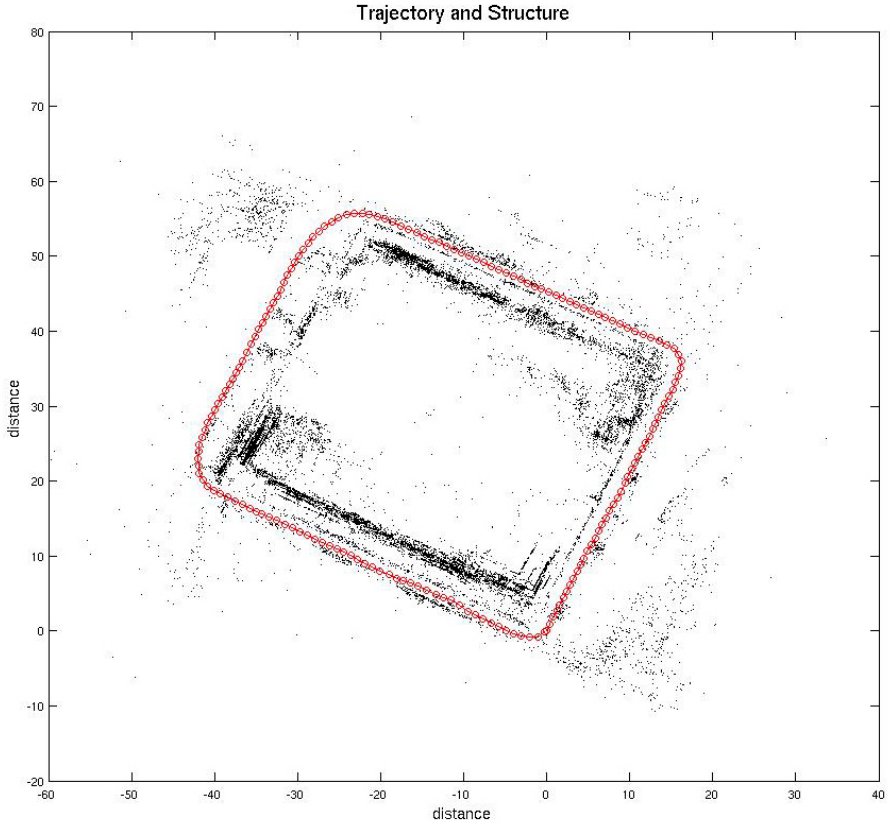


Fig. 5. Top: Estimated trajectory (red circles) and recovered point cloud for first set. Bottom: reconstructed 3D model. No global Bundle Adjustment was necessary.



Fig. 6. Sample images of second dataset





**Fig. 7.** Estimated trajectory (red circles) and recovered point cloud for second set. The camera started at the right-bottom corner moving towards the right side and ending in the same position. This was ensured using the same image for first and last position. No global Bundle Adjustment was necessary.

## 6 Conclusions

In this paper we present a closed form solution for the computation of the scale ratio between the translation directions of two consecutive image pairs. Our method requires only 1 2D-3D correspondence, though a least square solution is presented also in closed form. Combined with the 8-point algorithm and a local iterative refinement step we achieve accurate results with little drift over large trajectories. Our visual odometry algorithm is sufficiently accurate to be used for urban 3D reconstruction. We obtain 3D a reconstruction of a street without the need for global Bundle Adjustment. We also compare our method with a reference linear technique for the computation of a third camera pose using 6 2D-3D correspondences. Our method outperforms by a factor of 3 in

the computation of the rotation and a factor of 10 in the computation of the scale consistent translation. Additionally we provide a first order analysis of the propagation of the error, from the noise in the image features to the computation of the scale. Compared to alternative methods our approach takes full advantage of the larger number of image feature matches across consecutive frames. The 3-frame error propagation to the computation of the rotation and the translation direction is then avoided, effectively reducing the accumulation of error over time. The proposed algorithm is successfully applied to images recorded in a real urban environment for which we obtain accurate 3D reconstructions. In the future we plan to extend our work to take advantage of the error propagation in order to obtain a better estimate than a least square solution. This work and more will be available online at [www.fit3d.org](http://www.fit3d.org).

## References

1. Ameller, M.-A., Triggs, B., Quan, L.: Camera pose revisited - new linear algorithms. Rapport Interne - Equipe MOVI (2000)
2. Bujnak, M., Kukulova, Z., Pajdla, T.: A general solution to the p4p problem for camera with unknown focal length. In: CVPR (2008)
3. Dellaert, F., Seitz, S., Thorpe, C., Thrun, S.: Structure from motion without correspondence. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2000 ) (June 2000)
4. DeMenthon, D., Davis, L.S.: Exact and approximate solutions of the perspective-three-point problem. *IEEE Trans. Pattern Anal. Mach. Intell.* 14(11), 1100–1105 (1992)
5. Dubbelman, G., van der Mark, W., Groen, F.C.A.: Accurate and robust ego-motion estimation using expectation maximization. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (2008)
6. Durrant-Whyte, H., Bailey, T.: Simultaneous localisation and mapping (slam): Part i the essential algorithms. *Robotics and Automation Magazine* (2006)
7. Esteban, I., Dijk, J., Groen, F.: Fit3d toolbox: multiple view geometry and 3d reconstruction for matlab. In: International Symposium on Security and Defence Europe, SPIE (2010)
8. Eustice, R.M., Pizarro, O., Singh, H.: Visually augmented navigation for autonomous underwater vehicles. *IEEE J. Oceanic Eng.* (2007) (accepted, to appear)
9. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. pp. 726–740 (1987)
10. Hartley, R.I., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge (2000), ISBN: 0521623049
11. Horn, B.: *Recovering baseline and orientation from essential matrix* (1990)
12. Kanatani, K.: *Statistical Optimization for Geometric Computation*. Dover, New York (1996)
13. Lowe, D.: Distinctive image features from scale-invariant keypoints. *Int. J. of Computer Vision* (2004)
14. Moravec, H.: *Obstacle avoidance and navigation in the real world by a seeing robot rover* (1980)
15. Moreno Noguer, F., Lepetit, V., Fua, P.: Accurate non-iterative  $O(n)$  solution to the pnp problem. In: ICCV 2007, pp. 1–8 (2007)

16. Nister, D.: An efficient solution to the five-point relative pose problem. In: CVPR (2003)
17. Nister, D., Naroditsky, O., Bergen, J.: Visual odometry for ground vehicle applications. *Journal of Field Robotics* 23(1) (2006)
18. Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. *Int. J. Comput. Vision* 59(3), 207–232 (2004)
19. Scaramuzza, D., Fraundorfer, F., Siegwart, R.: Real-time monocular visual odometry for on-road vehicles with 1-point ransac. In: *IEEE International Conference on Robotics and Automation* (2009)
20. Scaramuzza, D., Siegwart, R.: Appearance guided monocular omnidirectional visual odometry for outdoor ground vehicles. *IEEE Transactions on Robotics*, Special Issue on Visual SLAM. Guest editors: José Neira, Andrew Davison, John J. Leonard (October 2008) (publication date)
21. Sur, F., Noury, N., Berger, M.-O.: Computing the uncertainty of the 8 point algorithm for fundamental matrix estimation. In: Everingham, M., Needham, C., Fraile, R. (eds.) *19th British Machine Vision Conference - BMVC 2008 Proceedings of the British Machine Vision Conference 2008, Leeds, Leeds United Kingdom (September 2008)*, <http://www.comp.leeds.ac.uk/bmvc2008/proceedings/papers/269.pdf>
22. Vedaldi, A.: An open implementation of the SIFT detector and descriptor. Technical Report 070012, UCLA CSD (2007)

# Recognizing 3D Human Motions Using Fuzzy Quantile Inference

Mehdi Khoury and Honghai Liu

Institute of Intelligent Systems and Robotics, Department of Creative Technologies,  
University of Portsmouth, Portsmouth, PO1 3QL, United Kingdom  
{mehdi.khoury, honghai.liu}@port.ac.uk

**Abstract.** Fuzzy Quantile Inference (FQI) is a novel method that builds a simple and efficient connective between probabilistic and fuzzy paradigms and allows the classification of noisy, imprecise and complex motions while using learning samples of suboptimal size. A comparative study focusing on the recognition of multiple stances from 3d motion capture data is conducted. Results show that, when put to the test with a dataset presenting challenges such as real biologically noisy” data, cross-gait differentials from one individual to another, and relatively high dimensionality (the skeletal representation has 57 degrees of freedom), FQI outperforms sixteen other known time-invariant classifiers.

## 1 Introduction

The process of behaviour understanding is usually performed by comparing observations to models inferred from examples using different learning algorithms. Such techniques presented in [1] and [21] can be used either in the context of template matching [3], state-spaces approaches [22], or semantic description [16]. Our application domain is focused on sport, and more precisely, boxing. We have discarded template matching as it is generally more susceptible to noise, variations of the time intervals of the movements, and is viewpoint dependent [21]. We are not interested in a pure semantic description as we need to analyze and evaluate a boxing motion in a relatively detailed way. We therefore focus on identifying static states during a motion (state-spaces approach). Conventionally, machine learning techniques in use for solving such problems vary from dynamic Time Warping [2], to Hidden Markov Models [15], Neural Networks [7], SVM [18], or variations of these techniques. This study presents FQI as a novel machine learning technique tested in the application domain of behaviour understanding, that is to say the recognition and description of actions and activities from the observation of human motions. It introduces a different method that allows us to build from learning samples fuzzy qualitative models corresponding to different states. An automated way to generate fuzzy membership function is proposed as an extension of Fuzzy Gaussian Inference [10]. It is applied to biologically “imprecise” human motion through the inference of underlying probability distributions behind the identified moves and their mapping to membership functions.

The first part of this paper defines and formalises the concept of FQI in the context of motion classification. Secondly, an experiment involving the comparison of FQI with 16 other algorithms over the classification of 3d motion capture data is conducted. Finally, future works are discussed in the conclusion.

## 2 Fuzzy Quantile Inference

To learn to recognize a stance, a model needs to be extracted (here a fuzzy membership function) for this stance from learning data. This stance is later identified during a motion by evaluating the membership score of the observed data with respect to the learned model. This chapter will first describe the model itself (the notion of fuzzy membership function). It will then describe the novel process by which this model is generated: Fuzzy Quantile Inference. Finally, it will show how the degree of membership of observed data to a given template is computed. But first, a brief description of the input data (human skeletal representation)in use is in order.

### 2.1 Human Skeletal Representation

We use the widely spread .BVH motion capture format [20] in which a human skeleton is formed of skeletal limbs linked by rotational joints. In practice, for every frame, our observed data takes the shape of a nineteen-by-three matrix describing ZXY Euler Angles for all nineteen joints in a simplified human skeletal representation (see [10] for more details). In other words, we seek to model the states of the 57 continuous variables (each between 0 and 360) that characterize a stance at any time.

### 2.2 Model Description: The Fuzzy Membership Function

The fuzzy linguistic approach introduced by Zadeh [23] allows us to associate a linguistic variable such as a “guard” stance with linguistic terms expressed by a fuzzy membership function. Using a trapezoid fuzzy-four-tuple  $(a, b, \alpha, \beta)$  which defines a function that returns a degree of membership in  $[0,1]$  (see equation 1) seems to be more interesting as there is a good compromise between precision and computational efficiency (compared with, for example, a triangular membership function).

$$\mu(x) = \begin{cases} 0 & x < a - \alpha \\ \alpha^{-1}(x - a + \alpha) & x \in [a - \alpha \quad a] \\ 1 & x \in [a \quad b] \\ \beta^{-1}(b + \beta - x) & x \in [b \quad b + \beta] \\ 0 & x > b + \beta \end{cases} \tag{1}$$

### 2.3 Model Generation Using Fuzzy Quantile Inference

Frames with a membership score equal to one for a given motion (e.g. “Guard”) are used as learning samples. The identification of these example data is made by a system similar to Reverse Rating [17], which is to say that, in our case, an expert (a human observer) is asked to do the following: identify a group of frames whose motion indicates a stance that possesses the degree 1.0 of membership in the fuzzy set “Guard”. Once these learning data are obtained, a fuzzy membership function can be generated. Many kinds of procedures for the automated generation of membership functions can be found in the literature. Fuzzy Clustering [9], Inductive Reasoning [11], Neural Networks [12] and Evolutionary Algorithms [19] have been used, among others, to build such functions. Estimation of S-Function from Histograms [5] has also been done, some

of it based on the optimization of fuzzy entropy [14]. So far, one downside of such techniques has been the difficulty to link the notion of fuzzy membership to the notion of probability distribution. One noticeable attempt to link both concepts in the generation of membership functions has been done by Frantti [6] in the context of mobile network engineering. Unfortunately, this approach is relatively limited as the minimum and maximum of the observed data are the absolute limits of the membership function. As a consequence, such a system ignores motions which are over the extremum of the learning range of the examples. This work presents a method that overcomes this problem by introducing a function that maps quantile functions to degrees of membership. Human motions are complex because they are unique. Nobody moves exactly the same way. Even the same person, when repeating the same motion, will not perform it in a strictly identical fashion. Despite of this, we seem to be able to recognize a motion when performed by different actors at various speeds and in very different conditions. From this observation, and the existence of the Central Limit Theorem, it is assumed that motions are somehow distributed. The characteristics of the distributions involved are unknown, but somehow, joints rotations, which are the most basic components of a motion, are assumed to be obeying some hidden probability density for a given motion. This assumption leads us to:

- Estimate the type of distribution underlying a given motion
- Map it to a representation system that is computationally efficient and suitable for noisy data (here, as explained previously, we choose trapezoid fuzzy membership functions).
- Evaluate this modelling process by learning and testing its accuracy and efficiency on real life motion capture data.

There exists many types of continuous distributions that can be applied to describe Euler angles. We do not pretend to list them all, but rather to find a systematic way to map best known distributions to one or several trapezoid fuzzy membership functions. We limit our analysis to distributions with only one mode (mapping saddle shaped distributions might be future work). Considering that a learning sample is of limited size, it represents a subset in the range of all possible joints rotations that characterize a given motion. As these are learning data, the joint rotations are expected to describe the target motion with an optimum qualitative certainty (with a membership score of 1). This has the following implication when building the fuzzy membership function: the extremum values of the learning sample will correspond to the lower and upper limits of the areas of membership 1 of the fuzzy membership function. Let the values of a variable  $X$  exposed in a learning sample be such that  $a \leq X \leq b$ . All values  $X$  in the interval  $[a, b]$  are given a membership score  $m = 1$ . Inversely, all values of  $X$  such that  $X$  belongs to  $]-\infty, a[ \cup ]b, +\infty[$  are given a membership score  $0 \leq m < 1$ . Therefore, a fuzzy membership function will be defined by the tuples  $\{a, b, \alpha, \beta\}$  where  $a$  and  $b$  are known and defined by the learning sample. There is now need to define the values of  $\alpha$  and  $\beta$  in order to complete the modeling process. The relative size  $s$  of the subset that constitutes the learning sample can be estimated as the proportion of the population whose value is in the interval  $[a, b]$ . This in fact corresponds to the average area under the distribution curve between  $a$  and  $b$ . This can be expressed as the difference of the cumulative distribution functions (or c.d.f.) of the variables  $b$  and  $a$  such that:

$$s = CDF(b) - CDF(a) = \int_a^b p(x)dx$$

Where  $p(x)$  is the probability density function (p.d.f.), for any two numbers  $a$  and  $b$  with  $a < b$ . From the parameter  $s$ , the lower and upper limits of the learning sample in number of standard deviations  $n$  on each side of the median of the distribution can be extracted. This number  $n$  can be extracted considering the quantile function also known as the inverse cumulative distribution function of the applied distribution. This gives us the distance ration between the maximum number of standard deviations covered by the fuzzy membership function and  $n$ . This allows to define the general shape of the fuzzy membership function in two steps: estimation based on a centralized mean of distribution, correction of estimation based on shift of central mean to sample mean.

**The estimation of a membership function based on a theoretical mean** is the first step towards building a model. Let  $s$  be a parameter such that:  $0 < s < 1$  describing the proportion of the population whose value is in the interval  $[a, b]$  of the min and max values of the learning sample. Let  $CDF^{-1}(x)$  be the quantile function also known as the inverse c.d.f of the applied distribution. Let  $CDF^{-1}(0.5)$  be the position of the median of the distribution. Depending on the distribution, let  $\mu$  be the estimated mean. Let  $a$ , the lowest value of the learning sample, be defined in term of the underlying distribution by  $a = CDF^{-1}(0.5 - (s/2))$ . Let  $b$ , the highest value of the learning sample, be defined in term of the underlying distribution by  $b = CDF^{-1}(0.5 + (s/2))$ . Let  $[\sigma_{min} \sigma_{max}]$  be the interval defining the values  $a - \alpha$  and  $b + \beta$  of the fuzzy membership function mapping the distribution. Then, as illustrated in Figure 1,  $\alpha_0$  and  $\beta_0$  are defined by:

$$\begin{cases} \alpha_0 = a - \sigma_{min} \\ \beta_0 = \sigma_{max} - b \end{cases} \tag{2}$$

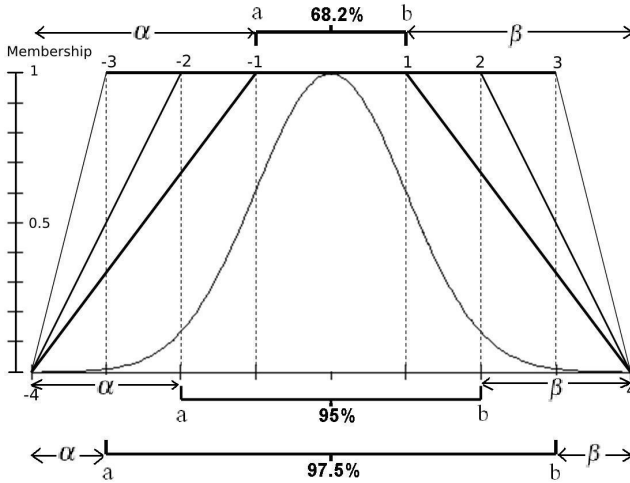
Normal, but also a Laplace, an Exponential and a Log-Normal distributions (see Figure 2) are considered. This first estimation is done assuming that the mean of the distribution overlaps with the mean of the learning sample. This calls for a reevaluation.

**The correction of the initial estimation** based on shift of central mean is the second step. By the Central Limit Theorem, the larger the learning sample the more likely its mean hints at the real position of the mean of the corresponding distribution. If the distribution has only one mode and is central symmetric:

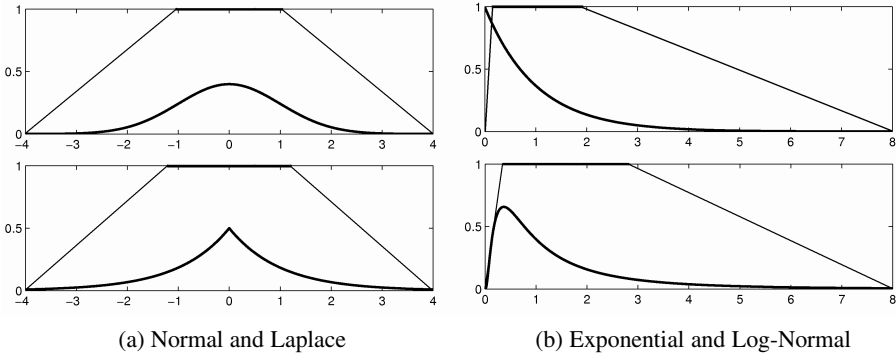
$$\begin{cases} \delta_\mu = (\mu_2 - a)/(b - a) \\ \alpha = (1 - \delta_\mu)(\alpha_0 + \beta_0) \\ \beta = (\delta_\mu)(\alpha_0 + \beta_0) \end{cases} \tag{3}$$

In the case of right-skewed distributions of the type Exponential or Log-Normal:

$$\begin{cases} \delta_\mu = (\mu_2 - \mu_1)/(b - \mu_1) \\ \alpha = \alpha_0 - (\delta_\mu \cdot \alpha_0) \\ \beta = \beta_0 + (\delta_\mu \cdot \alpha_0) \end{cases} \tag{4}$$



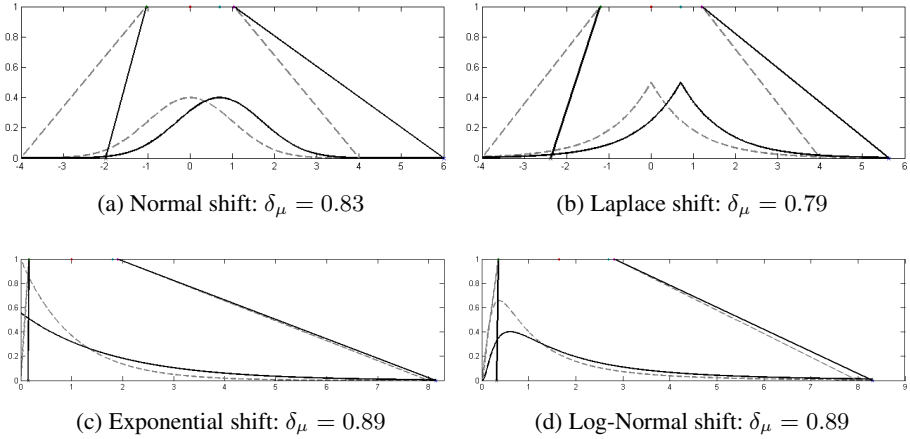
**Fig. 1.** Influence of  $s$  parameter on the shape of the Fuzzy Membership Function (FMF) for a Normal distribution



**Fig. 2.** Initial mapping of FMF from Normal and Laplace central symmetric and right-skewed distributions for  $s = 0.7$

To summarize the salient points of this method, considering the range and center of density of the learning sample, the shape of a fuzzy membership function will be defined by the following three factors. First, the maximum number of standard deviations covered by the fuzzy membership function. In the example of a normal distribution of standard deviation 1, the maximum range is approximated by assuming that it is  $n$  standard deviations away in both directions from the mid-point of the range of membership one. This will define the length of the base of the trapezoid shape. Secondly, depending on the value of the parameter  $s$ , a portion of the  $n$  standard deviations representing the total range will be allocated to the membership-one-range and the remaining part will





**Fig. 3.** Shifting distributions with  $s = 0.7$  depending on  $\delta_\mu$  (dashed lines are the shifted distributions)

be allocated to the lower membership degrees. This will define the length of the top part of the trapezoid shape (see Figure 1). Thirdly, the average of the means is extracted out of each learning sample. This will be used to shift the base of the trapezoid shape to either side in order to follow a shift of the distribution. This varies with the type of distribution (see equations 3 and 4).

**Membership Evaluation** is the last step. Observed data take the shape of a nineteen-by-three matrix describing ZXY Euler Angles for all nineteen joints. A “Guard” stance is defined by fuzzy membership functions that capture the Euler angles rotations defining a posture. One evaluates how close this matrix is from a “Guard” stance by calculating the degree of membership of every Euler Angle in every joint to its corresponding fuzzy-4-tuple. An average membership score is then computed. This approach could probably be improved in the near future by introducing weighted average for certain joints (for example, the position of the elbow might be more important than the position of the knee when in guard). In order to readjust the relative size of the samples, the membership scores  $s_i$  can be re-scaled by fine-tuning the thresholds  $t_i$  linked to each move  $i$  the following way :

$$s_i = (s_i - t_i) / (1 - t_i) \tag{5}$$

When a frame has a high membership score for several fuzzy sets, an order of preference of these sets can be established by comparing the Euclidian distance of the observed data to the centroid of each fuzzy set.

### 3 Comparing FQI with Other Classifiers: Experiment and Results

Boxing motions are chosen as they require precision, are well defined, and only require a small surface for the setting of a motion capture studio. They can also be reused in the

context of security surveillance. The experimental setting in use is essentially contrived as we are not using real fighting situations involving multiple opponents. Instead, participants are asked to perform modern boxing moves on their own, in a preordained and controlled fashion. Yet, this setting does have some natural aspects in that every similar motion performed by a human is unique in its own right. Motion-capture data from human participants are used as learning samples and testing samples. FQI is compared to sixteen time-invariant classification algorithms in order to recognize stances. These classifiers take exactly the same input data as FQI in order to provide a substantial ground for comparison. The description of the experiment starts with a review of the equipment, participants and procedures employed. Results are then exposed and discussed.

### 3.1 Apparatus

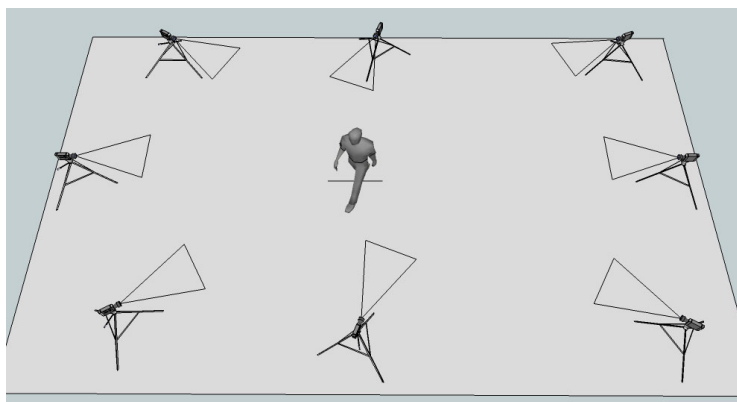
The motion capture data are obtained from a Vicon Motion Capture Studio with eight infra-red cameras. Each motion capture suits is set with a total of 49 optical passive reflective markers. The motion recognition is implemented in MATLAB on a standard PC with 2 Gigs of RAM. An additional MATLAB toolbox presented in [13] is also used for extracting Euler Angles from .BVH files.

### 3.2 Participants

Three male subjects, aged between 18 and 21, of light to medium-average size (167cm to 178cm) and weight (59 to 79kgs), all are competition level boxers at the national level. None of them present any abnormal gait. Optical Markers are placed in a similar way on each subject to ensure a consistent motion capture (see Figure 4). The motion capture Infra Red cameras are placed as shown in figure 5.



Fig. 4. Optical markers placement - front view



**Fig. 5.** Floor plan - the 3d motion capture studio

### 3.3 Procedure

The motion capture data are obtained from several subjects performing each boxing combination four times. There are twenty different boxing combinations, each separated by a guard stance. These are performed at two different speeds (medium-slow and medium fast). The boxing combinations are ordered sets of basic boxing stances. There are in total six precisely defined basic stances: Guard, Jab, Cross, Left Hook, Right Hook, Lower Left Hook. The rotational variations behind these stances are assumed to be normally distributed. These are well-known boxing stances and are accurately described in [4] (this description is valid for a right handed boxer - it should be reversed for a left-handed one).

Each participant goes through the following steps: on arrival, after a 10 min introduction, the participant is briefed on the experiment and given consent and information forms to read and sign. He is then putting the suit and captors on during the next 10 minutes. The next step which is the set up and calibration of the motion capture equipment takes around one hour. The data capture of different motions follows and is spread over 20 min. The participant then goes to the changing room and leaves.

A fuzzy membership function template corresponding to a specific stance is extracted from various samples. First, all three participants are employed to learn and to test how well the system recognizes the stances. Then, an evaluation is conducted to see how the system cope to learn from only two participants, and test how well it recognize stances from a third different participant. The accuracy of the system is examined when learning to recognize six different boxing stances simultaneously in a comparative work between FQI and other time invariant classifiers.

### 3.4 Results and Discussion

FQI is compared to algorithms that are implemented using the WEKA Machine Learning package presented in [8]. They represent diverse paradigms that can be roughly classified into seven types: Bayes, Function based, Nearest Neighbours, Tree based, Rule based, and Miscellaneous. The WEKA classifiers are used “as is” from the Weka

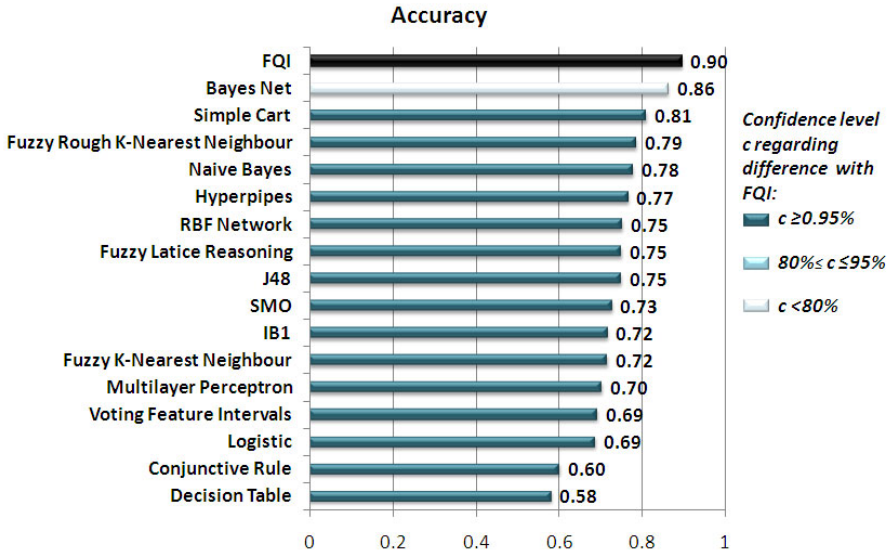


Fig. 6. Comparison of accuracy between FQI and 16 other classifiers

library using mostly default settings. Classifiers performances are evaluated using accuracy, and F-Score. Accuracy is defined as:

$$accuracy = \frac{tp + tn}{tp + fp + fn + tn} \tag{6}$$

where  $tp$  represents true positive rate,  $tn$  true negative rate,  $fp$  false positive rate, and  $fn$  false negative rate. It represents the proportion of true results (both true positives and true negatives) in the population. It is the most used empirical measure in the area of machine learning classifiers. At a first glance (see figure 6), results seem quite encouraging as FQI shows the best average accuracy of 90% over sixteen high performance machine learning classifying methods (figure 6). This difference in accuracy with FQI is significant at the level 0.05 on a two-tailed t-test for fifteen of these methods at the exception of Bayes Net (86%) which performs similarly and does not show a significant difference, even at the level 0.20. So far, if exclusively focusing on the accuracy, FQI and Bayes Net seem to be in first position, while the next three best methods are : SimpleCart (81%), Fuzzy Rough K-Nearest neighbours (79%) and Naive Bayes (78%). Beside FQI, the most accurate methods belong broadly to Bayes, decision trees and Fuzzy Rough Nearest Neighbours types. The worst performances belong to rule based classifiers - Conjunctive Rule (60%) and Decision Table (58%). They present a lost of accuracy of nearly 10% compared to the next best classifier. This gap could be explained by the difficulty for the bottom-up design to generate the sheer quantity of rules that could deal with real numbers in a noisy and biologically imprecise data set. When using accuracy alone, results seem to be in favor of FQI and Bayes Net. However, using only accuracy to measure the performance of the FQI classifier would only show a part

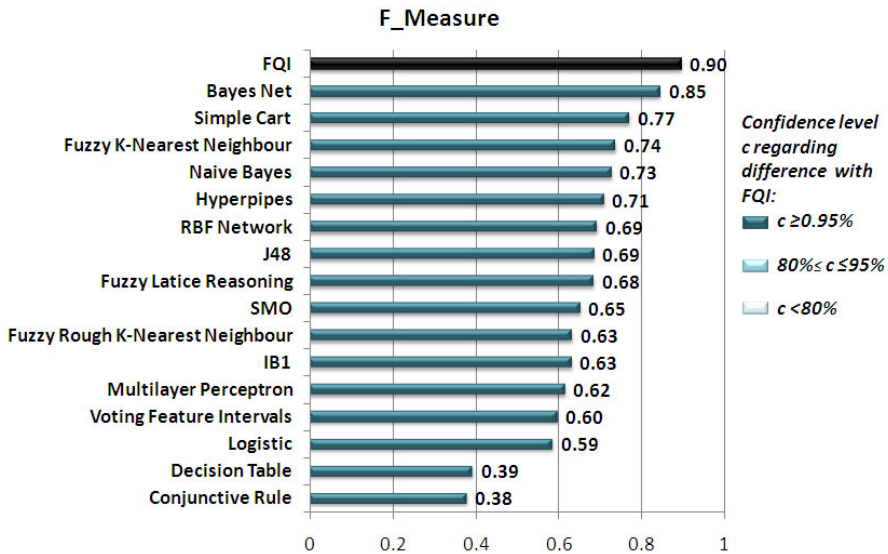


Fig. 7. Comparison of F-Measure between FQI and 16 other classifiers

of the whole picture. F-Measure or balanced F-Score (equation (7)) is a measure that combines Precision and Recall. It is the harmonic mean of precision and recall.

$$F - score = \frac{Precision \cdot Recall}{Precision + Recall} \tag{7}$$

FQI shows the best average F-score of 90% over the sixteen other classifiers (figure (6)). This difference in F-score with FQI is significant at the level 0.05 on a two-tailed t-test for all of these methods. The next four best classifiers are Bayes Net (85%), Simple-Cart (77%), Fuzzy Rough K-Nearest neighbours (74%) and Naive Bayes (73%). These results confirm the overall dominance of FQI so far.

FQI seems to compare favorably with the considered time invariant techniques when classifying static stances. Results on this particular data set show that it ends up with the best Accuracy, and F-Score. These positive results suggest that FQI seems to emulate certain successful traits present in other classifiers such as: the ability to decouple conditional feature distributions into one-dimensional sets, the capacity to link these features by combining their memberships, the ability to deal with imprecision and overlapping classes scores, and the capacity to deal with partial knowledge by taking into account the degree of incompleteness of the learning sample when mapping it to an underlying distribution.

## 4 Conclusion

The novelty of FQI comes from the fact that it builds a simple and efficient connective between probabilistic and fuzzy paradigms that allows the classification of noisy,

imprecise and complex motions while using learning samples of sub-optimal size. Results show that, when put to the test with a dataset presenting challenges such as biologically noisy” data, cross-gait differentials from one individual to another, and high dimensionality caused by the complexity of the skeletal representation (57 degrees of freedom), FQI outperforms 16 other known time invariant classifiers. This, somehow, gives FQI stronger credentials as a contender in the field of motion recognition. However, if FQI is relatively good at identifying static stances, it cannot so far take into account previous poses, or contextual information. Considering the similarity of some static poses that belong to different moves, this unavoidably causes mistakes within the classification results. There is therefore need to introduce a time dimension and contextual information to refine the results, and also need make the system cope with missing features that could result from visual occlusion.

## Acknowledgements

The project is funded by EPSRC Industrial CASE studentship and MM2G Ltd under grant No 07002034. Many thanks to Portsmouth University Boxing Club and to the Motion capture Team: A. Counsell, G. Samuel, O. Seymour, I. Sedgebeer, D. McNab, D. Shipway and M. Mitrofanov.

## References

1. Aggarwal, J.K., Cai, Q., Liao, W., Sabata, B.: Articulated and elastic non-rigid motion: A review. In: Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects, pp. 2–14 (1994)
2. Bobick, A.F., Wilson, A.D.: A state based technique for the summarization and recognition of gesture. In: International Conference on Computer Vision, pp. 382–388 (1995)
3. Chan, C.S., Liu, H., Brown, D.J.: Recognition of human motion from qualitative normalised templates. *Journal of Intelligent and Robotic Systems* 48(1), 79–95 (2007)
4. Cokes, C.: *The Complete Book of Boxing for Fighters and Fight Fans*. Etc. Publications (1980)
5. Devi, B.B., Sarma, V.V.S.: Estimation of fuzzy memberships from histograms. *Inf. Sci.* 35(1), 43–59 (1985)
6. Frantti, T.: Timing of fuzzy membership functions from data. Academic Dissertation, University of Oulu, Finland (July 2001)
7. Guo, Y., Xu, G., Tsuji, S.: Understanding human motion patterns. In: International Conference on Pattern Recognition, pp. B325–B329 (1994)
8. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.: The weka data mining software: An update. *ACM SIGKDD Explorations Newsletter* 11(1), 10–18 (2009)
9. Iokibe, T.: A method for automatic rule and membership function generation by discretionary fuzzy performance function and its application to a practical system. In: Proceedings of the First International Joint Conference of the North American Fuzzy Information Processing Society Biannual Conference, pp. 363–364 (1994)
10. Khoury, M., Liu, H.: Fuzzy qualitative gaussian inference: Finding hidden probability distributions using fuzzy membership functions. In: IEEE Workshop on Robotic Intelligence in Informationally Structured Space, RiiSS 2009 (2009)

11. Kim, C., Russell, B.: Automatic generation of membership function and fuzzy rule using inductive reasoning. In: *Industrial Fuzzy Control and Intelligent Systems*, pp. 93–96 (1993)
12. Kim, J., Seo, J., Kim, G.: Estimating membership functions in a fuzzy network model for part-of-speech tagging. *Journal of Intelligent and Fuzzy Systems* 4, 309–320 (1996)
13. Lawrence, N.D.: *Mocap toolbox for matlab* (2007), <http://www.cs.man.ac.uk/~neill/mocap/>
14. Nieradka, G., Butkiewicz, B.S.: A method for automatic membership function estimation based on fuzzy measures. In: Melin, P., Castillo, O., Aguilar, L.T., Kacprzyk, J., Pedrycz, W. (eds.) *IFSA 2007. LNCS (LNAI)*, vol. 4529, pp. 451–460. Springer, Heidelberg (2007)
15. Pentland, A.P., Oliver, N., Brand, M.: Coupled hidden markov models for complex action recognition. In: *Massachusetts Institute of Technology, Media Lab* (1996)
16. Remagnino, P., Tan, T.N., Baker, K.D.: Agent orientated annotation in model based visual surveillance. In: *International Conference on Computer Vision*, pp. 857–862 (1998)
17. Sanghi, S.: Determining membership function values to optimize retrieval in a fuzzy relational database. In: *Proceedings of the 2006 ACM SE Conference*, vol. 1, pp. 537–542 (2006)
18. Schüldt, C., Laptev, I., Caputo, B.: Recognizing human actions: A local svm approach. In: *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR 2004)*, vol. 3, pp. 32–36 (2004)
19. Simon, D.: Hinfinity estimation for fuzzy membership function optimization. *Int. J. Approx. Reasoning* 40(3), 224–242 (2005)
20. Thingvold, J.: *Biovision bvh format* (1999) <http://www.cs.wisc.edu/graphics/Courses/cs-838-1999/Jeff>
21. Wang, L., Hu, W., Tan, T.: Recent developments in human motion analysis. *Pattern Recognition* 36(3), 585–601 (2003)
22. Yamato, J., Ohya, J., Ishii, K.: Recognizing human action in time-sequential images using hidden markov model. In: *IEEE Computer Vision and Pattern Recognition*, pp. 379–385 (1992)
23. Zadeh, L.: Fuzzy sets. *Information and Control* 8, 338–353 (1986)

# Evaluating the Fuzzy Coverage Model for 3D Multi-camera Network Applications

Aaron Mavrincac, Jose Luis Alarcon Herrera, and Xiang Chen

University of Windsor, Windsor, Ontario, Canada  
{mavrinc1, alarconj, xchen}@uwindsor.ca

**Abstract.** An intuitive three-dimensional task-oriented coverage model for 3D multi-camera networks based on fuzzy sets is presented. The model captures the vagueness inherent in the concept of visual coverage, with a specific target of the feature detection and matching task. The coverage degree predicted by the model is validated against various multi-camera network configurations using the SIFT feature detection and description algorithm.

## 1 Introduction

Three-dimensional multi-camera networks (that is, camera networks with three-dimensional sensing capabilities) cover a major subset of camera network applications. A fuzzy representation lends itself well to modeling the vague concept of task-specific coverage in such a network. Spatial points with associated direction form the universal set of observable information, and assigning to these a degree of membership in the set of covered points yields a quantitative metric for coverage.

In other work, we shall present a full theoretical derivation of the *fuzzy coverage model*. It is a comprehensive three-dimensional coverage model for cameras and multi-camera networks, based on fuzzy sets, which is derived from well-studied parameters of the standard imaging system model (obtained from internal and external calibration), and can be tuned for specific tasks via a small number of intuitive application parameters. The model takes its inspiration, in part, from work by Cowan and Kovesi [1] and others. Here, we focus on describing and evaluating the 3D multi-camera network coverage model, and dispense with much of the background theory as it is beyond our scope.

A major task requirement for most 3D applications is the detection and matching of local features. Modern popular algorithms such as SIFT [2], SURF [3], MSER [4], and others allow for fairly reliable sparse feature matching between relatively widely separated views – and, therefore, wide-baseline stereo reconstruction. In order to evaluate the coverage of arbitrary scene features, the nature of which (including planar orientations) are not known a priori, directional coverage must be modeled in the scene. As will be shown, the fuzzy coverage model captures this aspect of coverage quite well.

The remainder of this paper is organized thus. Section 2 presents some of the underlying theory and conventions used in our model. Section 3 describes the



model specifically for 3D multi-camera networks, with brief insights into the fundamental computer vision theory. Experimental results for the feature detection and matching task are presented in Section 4. Finally, concluding remarks are given in Section 5.

## 2 Definitions and Conventions

### 2.1 Geometric Conventions

When dealing with a three-dimensional Euclidean space, we express positions and orientations in a right-handed Cartesian coordinate system, with the axes of its basis denoted  $x$ ,  $y$ , and  $z$ , and fixed-axis rotation angles about these axes denoted  $\theta$ ,  $\phi$ , and  $\psi$ , respectively.

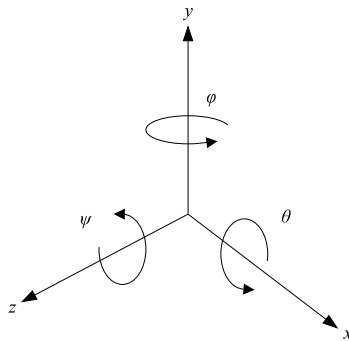


Fig. 1. Axes and Rotations

Naturally, for the spatial coordinates,  $(x, y, z) \in \mathbb{R}^3$ . Similarly, for the rotation angles,  $(\theta, \phi, \psi) \in [0, 2\pi]^3$ ; for every equivalence class on angles  $\bar{\theta}_{2\pi}$  we consider the value  $\theta \in [0, 2\pi)$ .

For vector direction, we also employ a two-angle orientation defined by inclination angle  $\rho \in [0, 2\pi)$ , measured from the  $z$ -axis zenith, and azimuth angle  $\eta \in [0, 2\pi)$ , measured from the  $x$ -axis in the  $x$ - $y$  plane (in the direction of  $\psi$ ). Since, for  $0 \leq e \leq \pi$ ,  $(\pi - e, \eta) \sim (\pi + e, \eta + \pi)$ , we generally restrict  $\rho$  to  $[0, \pi]$  where possible. Note also that  $(0, \eta_1) \sim (0, \eta_2)$  and  $(\pi, \eta_1) \sim (\pi, \eta_2)$  for any  $\eta_1$  and  $\eta_2$ .

**Definition 1.** *The directional space  $\mathbb{D} = \mathbb{R}^3 \times [0, \pi] \times [0, 2\pi)$  consists of three-dimensional Euclidean space plus direction, with elements of the form  $(x, y, z, \rho, \eta)$ .*

For convenience, we denote the spatial component  $\mathbf{p}_s = (x, y, z)$  and the directional component  $\mathbf{p}_d = (\rho, \eta)$ .

A standard 3D pose  $P : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ , consisting of rotation matrix  $\mathbf{R}$  and translation vector  $\mathbf{T}$ , may be applied to  $\mathbf{p} \in \mathbb{D}$ . The spatial component is transformed as usual, i.e.  $P(\mathbf{p}_s) = \mathbf{R}\mathbf{p}_s + \mathbf{T}$ . The direction component is transformed as follows. If  $\mathbf{d}$  is the unit vector in the direction of  $\mathbf{p}_d$ , then  $P(\mathbf{p}_d) =$

$(\arccos(\mathbf{Rd}_z), \arctan 2(\mathbf{Rd}_y, \mathbf{Rd}_x))$ , where  $\arctan 2$  is an arctangent variant returning the angle from the  $x$ -axis to a vector in the full range  $[0, 2\pi)$ .

### 2.2 Fuzzy Sets

A fuzzy set [5] is a pair  $(S, \mu)$ , where  $S$  is a set (called the *universal set*) and  $\mu : S \rightarrow [0, 1]$  is a membership function indicating the grade of membership of elements in  $S$  to the fuzzy set. For a fuzzy set  $A = (S_A, \mu_A)$ , the set  $\text{supp}(A) = \{x \in S_A | \mu_A(x) > 0\}$  is called the *support* of  $A$ , and the set  $\text{kern}(A) = \{x \in S_A | \mu_A(x) = 1\}$  is called its *kernel*.

The standard *fuzzy union* operation is defined for fuzzy sets  $A$  and  $B$  as  $A \cup B = (S_A \cup S_B, \mu_{A \cup B})$ , where  $\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$  for  $x \in S_A \cup S_B$ . Similarly, the standard *fuzzy intersection* operation is defined as  $A \cap B = (S_A \cap S_B, \mu_{A \cap B})$ , where  $\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$  for  $x \in S_A \cap S_B$ . Other t-conorms and t-norms yield different union and intersection operations [6]; in particular, we also use the algebraic product t-norm for intersection, for which  $\mu_{A \cap B}(x) = \mu_A(x) \cdot \mu_B(x)$ .

In modeling the coverage of cameras and multi-camera networks, we deal with fuzzy subsets of  $\mathbb{R}^3$  and  $\mathbb{D}$ .

A continuous fuzzy set  $A$  has a discrete counterpart, denoted  $\dot{A}$ , which is formed by sampling a finite number of elements from  $A$  such that  $\mu_{\dot{A}}(\mathbf{p}) = \mu_A(\mathbf{p})$  for all  $\mathbf{p} \in S_{\dot{A}}$ . We adopt this notation for all discrete fuzzy sets. A discrete fuzzy set  $\dot{A}$  has *scalar cardinality*  $|\dot{A}| = \sum_{x \in S_{\dot{A}}} \mu_{\dot{A}}(x)$ .

## 3 The Fuzzy Coverage Model

### 3.1 Single Camera

Consider a fuzzy subset  $C \in \mathbb{D}$ , with  $\mathbb{R}^3$  in the camera frame basis, where membership grade  $\mu_C(\mathbf{p})$  indicates “good imaging” at spatial point  $\mathbf{p}_s$  from direction  $\mathbf{p}_d$ . In other words,  $C$  is the fuzzy set of directional points (in the camera frame) which are *covered*.

Our realization of  $C$  is entirely determined from nine intrinsic parameters of the imaging system ( $A, f, s_u, s_v, o_u, o_v, w, h,$  and  $z_S$ ) and five application parameters ( $\gamma, R_1, R_2, c_{\max},$  and  $\zeta$ ) for a total of 14 parameters. The intrinsic parameters are normally obtained from camera specifications and calibration, and the application parameters are intuitive to set based on task requirements. Four individual factors – visibility, resolution, focus, and direction – yield component fuzzy subsets of  $\mathbb{D}$ . These components are combined via algebraic product fuzzy intersection:

$$C = C_V \cap C_R \cap C_F \cap C_D \tag{1}$$

Conveniently, if it is not desirable to account for a component in a certain application, it can simply be left out of (1). In the particular case where directional visibility is not relevant (thus  $C = C_V \cap C_R \cap C_F$ ),  $C$  can be treated as a fuzzy subset of  $\mathbb{R}^3$  rather than of  $\mathbb{D}$  for practical purposes, since, as will be seen, the directional components have no effect.

**Visibility.** Visibility is a bivalent condition depending whether  $\mathbf{p}$  is within the field of view, a function of  $x$ ,  $y$ , and  $z$ . However, points near the edges may not be considered “fully visible” for an application’s purposes, so a general-purpose application parameter  $\gamma$  is introduced to fuzzify visibility into a trapezoidal membership function. The value of  $\gamma$  is simply the number of pixels in from the edge of the image where features are considered to be fully within the field of view.

$C_V$  has  $\text{kern}(C_V) = \{(x, y, z, \rho, \eta) \in \mathbb{D} \mid -\sin(\alpha_{hl}) + \gamma_h \leq x \leq \sin(\alpha_{hr}) - \gamma_h, -\sin(\alpha_{vl}) + \gamma_v \leq y \leq \sin(\alpha_{vr}) - \gamma_v, z > 0\}$  and  $\text{supp}(C_V) = \{(x, y, z, \rho, \eta) \in \mathbb{D} \mid -\sin(\alpha_{hl}) \leq x \leq \sin(\alpha_{hr}), -\sin(\alpha_{vl}) \leq y \leq \sin(\alpha_{vr}), z > 0\}$ , where  $\gamma_h$  and  $\gamma_v$  are calculated from  $\gamma$  as follows:

$$\gamma_h = \frac{2\gamma}{w} \sin\left(\frac{\alpha_h}{2}\right) \tag{2}$$

$$\gamma_v = \frac{2\gamma}{h} \sin\left(\frac{\alpha_v}{2}\right) \tag{3}$$

Here,  $\alpha_h$  and  $\alpha_v$  are the horizontal and vertical angles of view (computed from calibration), with  $\alpha_{hl}$ ,  $\alpha_{hr}$ ,  $\alpha_{vl}$ , and  $\alpha_{vr}$  being the half-angles measured from the principal axis.

**Resolution.** Resolution is inversely proportional to the  $z$ -coordinate of  $\mathbf{p}$ .  $C_R$  has  $\text{kern}(C_R) = \{(x, y, z, \rho, \eta) \in \mathbb{D} \mid 0 \leq z \leq z_1\}$  and  $\text{supp}(C_R) = \{(x, y, z, \rho, \eta) \in \mathbb{D} \mid 0 \leq z \leq z_2\}$ .

The values of  $z_1$  and  $z_2$  are computed from application parameters  $R_1$  and  $R_2$  as follows:

$$z_R = \frac{1}{R} \min \left[ \frac{w}{2 \sin(\alpha_h/2)}, \frac{h}{2 \sin(\alpha_v/2)} \right] \tag{4}$$

where  $w$  and  $h$  are the pixel width and height of the image, respectively.  $R_1$  is the minimum pixel resolution at which the application is not negatively affected by the resolution, and  $R_2$  is the absolute minimum pixel resolution beyond which the application does not produce acceptable results.

**Focus.** Focus is also a function of  $z$ .  $C_F$  has  $\text{kern}(C_F) = \{(x, y, z, \rho, \eta) \in \mathbb{D} \mid z_{\triangleleft} \leq z \leq z_{\triangleright}\}$  and  $\text{supp}(C_F) = \{(x, y, z, \rho, \eta) \in \mathbb{D} \mid z_n \leq z \leq z_f\}$ .

For  $z \in [z_{\triangleleft}, z_{\triangleright}]$ , the digital image is effectively in perfect focus; the values are calculated as follows:

$$z_{\triangleleft} = \frac{Afz_S}{Af + s(z_S - f)} \quad z_{\triangleright} = \frac{Afz_S}{Af - s(z_S - f)} \tag{5}$$

where  $A$  is the aperture diameter,  $f$  is the focal length,  $z_S$  is the focus distance, and  $s$  is the real size of one pixel.

The values of  $z_n$  and  $z_f$  are the near and far limits of depth of field for the maximum circle of confusion diameter, application parameter  $c_{\max}$ :

$$z_n = \frac{Afz_S}{Af + c_{\max}(z_S - f)} \quad z_f = \frac{Afz_S}{Af - c_{\max}(z_S - f)} \tag{6}$$

**Direction.** Directional visibility is a function of inclination angle  $\rho$ . An application parameter  $\zeta$  is introduced to fuzzify directional visibility into a trapezoidal membership function, which can be used to account for the quality of coverage of an object face dependent on viewing angle, or for the ability to match features.

The fuzzy subset for direction  $C_D$  has  $\text{kern}(C_D) = \{(x, y, z, \rho, \eta) \in \mathbb{D} \mid \rho \geq \pi/2 + \Theta + \zeta\}$  and  $\text{supp}(C_D) = \{(x, y, z, \rho, \eta) \in \mathbb{D} \mid \rho \geq \pi/2 + \Theta\}$ , where

$$\Theta = \left( \frac{y}{r} \sin \eta + \frac{x}{r} \cos \eta \right) \arctan \left( \frac{r}{z} \right) \tag{7}$$

with  $r = \sqrt{x^2 + y^2}$  and for  $z > 0$ .

The value of  $\zeta$  is in the range  $[0, \pi/2]$ , and reflects the angle of the surface normal of a feature, relative to the principal axis, at which performance for a given task begins to degrade.

### 3.2 3D Multi-camera Network

In a 3D multi-camera network, coverage depends on at least two cameras imaging the same point (including direction). Typically, applications perform 3D reconstruction or other pairwise processing of shared scene.

**In-Scene Single-Camera Model.** The *in-scene model* for a single camera,  $C^s$ , is simply the camera model  $C$  transformed to the world frame. Thus,  $C^s$  has six additional parameters – the pose, or extrinsic parameters, of the camera, defined by  $x, y, z, \theta, \phi$ , and  $\psi$  – for a total of 20 parameters.

**Adding the Occlusion Constraint.** In order to introduce the additional constraint of scene occlusion, the scene model  $S$  is required. This consists of a series of opaque plane segments (surfaces of static scene objects). A spatial point  $\mathbf{p}_s$  is occluded iff there exists a unique point of intersection  $\mathbf{p}_i \neq \mathbf{p}_s$  between the line segment from  $\mathbf{p}_s$  to the camera’s principal point  $\mathbf{p}_p$  and any opaque plane segment.

$$\mathbf{p}_i = \mathbf{p}_s + (\mathbf{p}_p - \mathbf{p}_s)t \tag{8}$$

where  $t$  is found, given three non-collinear points  $P_0, P_1$ , and  $P_2$  in a plane  $P \in S$ , from

$$\begin{bmatrix} t \\ u \\ v \end{bmatrix} = \begin{bmatrix} \mathbf{p}_{sx} - \mathbf{p}_{px} & P_{1x} - P_{0x} & P_{2x} - P_{0x} \\ \mathbf{p}_{sy} - \mathbf{p}_{py} & P_{1y} - P_{0y} & P_{2y} - P_{0y} \\ \mathbf{p}_{sz} - \mathbf{p}_{pz} & P_{1z} - P_{0z} & P_{2z} - P_{0z} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{p}_{sx} - P_{0x} \\ \mathbf{p}_{sy} - P_{0y} \\ \mathbf{p}_{sz} - P_{0z} \end{bmatrix} \tag{9}$$

The in-scene model with occlusion is denoted  $C^o$ .

**Pairwise Coverage.** Because the 3D multi-camera network performs pairwise processing, we must consider the coverage models of *pairs* of cameras. The pairwise coverage model for cameras  $k$  and  $l$  is a fuzzy subset  $C_{kl}^o = C_k^o \cap C_l^o$ .

Network coverage for a 3D multi-camera network can be modeled as a fuzzy subset  $C_N = \bigcup_{k,l \in N^2} C_{kl}^o$ .

## 4 Experimental Application

### 4.1 Feature Matching

Many multi-camera network applications, particularly those involving 3D reconstruction, rely heavily on matching local image features. A variety of feature detection and description algorithms exist, but a major common limitation is degradation of performance over large rotational transformations of the viewpoint [7,8,9,10].

Conceptually, the membership degree of a  $\mathbb{D}$  vector in  $C_{kl}^o$  will depend heavily on the rotation component of  $P_{kl}$ : the larger the angle between the principal axes of  $k$  and  $l$ , the smaller the range of directions covered in common. The  $\zeta$  application parameter is tuned according to the repeatability of a feature detector over such rotation (experimentally, in our case).

In our experiments, we test the detection and matching performance of the popular SIFT [2] algorithm for local feature detection and description. A total of 26 camera views of a scene of known structure containing 56 features with known 3D positions are used to validate the detection and matching performance predicted by the fuzzy coverage model.

### 4.2 Apparatus

**Software Implementation.** The fuzzy coverage model and fuzzy vision graph have been implemented in an object-oriented software stack using Python [11]. First, we have developed FuzzPy [12], a generic open-source Python library for fuzzy sets and graphs. Using this functionality, we have developed various classes for the fuzzy sets used in the model. The `Camera` class, initialized with the 14 model parameters, returns the  $\mu$ -value of any spatial point in camera coordinates using continuous trapezoidal fuzzy sets to implement  $C$ . The `MultiCameraSimple` and `MultiCamera3D` classes build their respective discrete fuzzy sets from `Camera` objects and a supplied scene model  $S$ . Coverage performance  $m$  can be estimated given a discrete fuzzy subset  $\hat{D} \in \mathbb{D}$ .

**Cameras.** Prosilica EC-1350 cameras, with a sensor resolution of  $1360 \times 1024$  pixels and square pixel size of  $4.65 \mu\text{m}$ , are employed. These are fitted with Computar M3Z1228C-MP manual varifocal lenses, with a focal length range of 12 mm to 36 mm and a maximum aperture ratio of 1:2.8. Calibration and image acquisition are performed using HALCON [13]. The SIFT implementation from VLFeat [14] is used for feature detection and matching.

### 4.3 Experimental Results

An experimentally determined value of  $\zeta = 0.3$  is used for the application parameter for directional visibility (other application parameters are  $\gamma = 20$ ,  $R_1 = 3.0$ ,  $R_2 = 0.5$ ,  $c_{\text{max}} = 0.0048$ ).

A four-camera network is used to evaluate the model in this section. One image is taken from each camera, then SIFT is used to find the features in all

the images and the matching points in all the pairs for the network. The scene is then simulated and the results are shown in Tables 1 through 6. Figure 2 shows the four images used for this setup, and Figure 3 shows the simulated model used to visualize the predicted performance of this camera setup.

When SIFT is used the result is a bivalent condition that indicates whether the feature, or point in this case, was detected by the algorithm in both images and consequently matched for the given pair. While 1 indicates that the feature was found in both images 0 represents two possibilities: the feature was not detected in any image or it was only detected in one image.

For pair 6 no matches were found, that result is reflected in the model since there are no points with  $\mu > 0$ .

**Table 1.** Camera Network Pair 1

Points	$P_{06}$	$P_{13}$	$P_{14}$	$P_{15}$	$P_{38}$	$P_{39}$	$P_{46}$	$P_{47}$
SIFT	1	1	1	1	1	1	1	1
Model	0.72	0.00	0.51	0.00	0.70	0.00	0.50	0.58

**Table 2.** Camera Network Pair 2

Points	$P_{06}$	$P_{13}$	$P_{14}$	$P_{15}$	$P_{38}$	$P_{39}$	$P_{46}$	$P_{47}$
SIFT	1	1	1	1	1	1	1	1
Model	0.94	0.71	0.71	0.24	0.91	0.92	0.82	0.81

**Table 3.** Camera Network Pair 3

Points	$P_{01}$	$P_{02}$	$P_{05}$	$P_{06}$	$P_{09}$	$P_{13}$	$P_{14}$	$P_{15}$	$P_{16}$	$P_{17}$
SIFT	0	0	1	1	0	1	1	1	0	0
Model	0.77	0.81	0.82	0.72	0.28	0.00	0.51	0.00	0.37	0.47

**Table 4.** Camera Network Pair 3 continued

Points	$P_{18}$	$P_{26}$	$P_{27}$	$P_{28}$	$P_{37}$	$P_{38}$	$P_{39}$	$P_{46}$	$P_{47}$	$P_{48}$
SIFT	0	1	1	1	1	1	1	1	1	1
Model	0.46	0.00	0.76	0.68	0.76	0.70	0.00	0.50	0.58	0.51

**Table 5.** Camera Network Pair 4

Points	$P_{01}$	$P_{02}$	$P_{09}$	$P_{16}$	$P_{17}$	$P_{18}$	$P_{27}$	$P_{28}$	$P_{40}$	$P_{48}$
SIFT	0	0	1	0	1	1	1	1	0	1
Model	0.77	0.81	0.82	0.37	0.57	0.65	0.62	0.68	0.63	0.51

**Table 6.** Camera Network Pair 5

Points	$P_{01}$	$P_{02}$	$P_{09}$	$P_{16}$	$P_{17}$	$P_{18}$	$P_{25}$	$P_{26}$	$P_{27}$	$P_{28}$	$P_{48}$	$P_{55}$
SIFT	1	1	0	1	0	0	0	0	1	1	1	1
Model	0.85	0.81	0.28	0.42	0.47	0.46	0.54	0.55	0.62	0.70	0.91	0.86

Tables 4 through 6 show that the points with a SIFT condition of 1 also have a  $\mu > .5$  for this camera network, this is considered a good prediction from the camera model, however there are some points that have a SIFT condition of 1 and a  $\mu = 0$ , regardless these points are considered as false positives because they do not appear as points in the images and SIFT does not look for any specific shape.

Of the total of points that are in the scene 37.5% appear in the model with a  $\mu > .5$  and 33.92% appear with a  $\mu > .5$  and were also detected by the SIFT implementation.

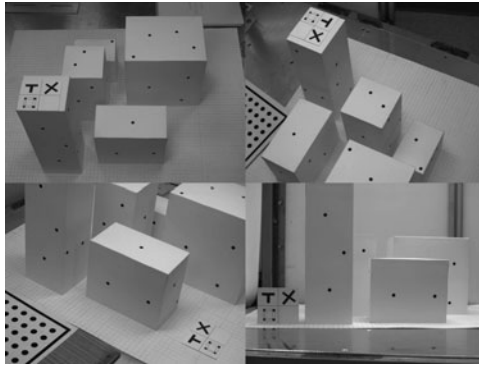


Fig. 2. Views for Camera Network 1

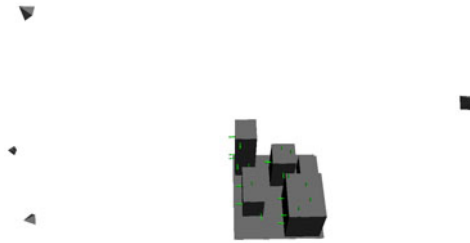
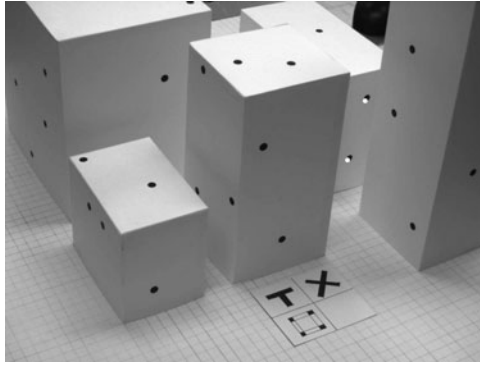


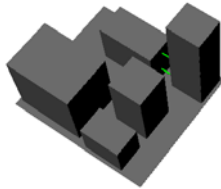
Fig. 3.  $M^0$  for Camera Network 1

Figure 4 shows the matched points for a particular pair of images from a different camera setup. Each camera covers a large amount of features; however, due to the pose of each camera the amount of features covered in common is small. This is reflected in the model shown in Fig 5.

In Figure 5 the points being matched by the SIFT implementation are  $P_{44}$  and  $P_{45}$ ; the  $\mu$  values for these points are  $\mu_{P_{44}} = 0.67$  and  $\mu_{P_{45}} = 0.59$ .



**Fig. 4.** Matched points for a pair of cameras



**Fig. 5.** Model for cameras in Fig. 4

## 5 Conclusions

The fuzzy coverage model accurately encapsulates information about a 3D multi-camera network's predicted performance on the ubiquitous task of feature detection and matching.

## References

1. Cowan, C.K., Kovese, P.D.: Automatic Sensor Placement from Vision Task Requirements. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 10(3), 407–416 (1988)
2. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. *Intl. Jrnl. of Computer Vision* 60(2), 91–110 (2004)



3. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. In: Proc. 9th European Conf. on Computer Vision, pp. 404–417 (2006)
4. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In: Proc. British Machine Vision Conf., pp. 384–393 (2002)
5. Zadeh, L.A.: Fuzzy Sets. *Information and Control* 8(3), 338–353 (1965)
6. Klir, G.J., Yuan, B.: *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice-Hall, Englewood Cliffs (1995)
7. Baumberg, A.: Reliable Feature Matching Across Widely Separated Views. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1774–1781 (2000)
8. Fraundorfer, F., Bischof, H.: Evaluation of Local Detectors on Non-Planar Scenes. In: Proc. 28th Wkshp. of the Austrian Association for Pattern Recognition, pp. 125–132 (2004)
9. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A Comparison of Affine Region Detectors. *Intl. Jnl. of Computer Vision* 65(1), 43–72 (2006)
10. Moreels, P., Perona, P.: Evaluation of Features Detectors and Descriptors Based on 3D Objects. *Intl. Jnl. of Computer Vision* 73(3), 263–284 (2007)
11. Python Software Foundation: Python
12. Mavrinac, A., Spriet, X.: FuzzPy
13. MVTec Software GmbH: HALCON
14. Vedaldi, A., Fulkerson, B.: VLFeat: An Open and Portable Library of Computer Vision Algorithms (2008)

# Emotion Recognition in Facial Image Sequences Using a Combination of AAM with FACS and DBN\*

Kwang-Eun Ko and Kwee-Bo Sim

School of Electrical and Electronics Engineering, Chung-Ang University,  
221, Heuksuk-Dong, Dongjak-Ku, Seoul, Korea  
kbsim@cau.ac.kr

**Abstract.** If we want to recognize the human's emotion via the face-to-face interaction, first of all, we need to extract the emotional features from the facial image and recognize the emotional states. Our facial emotional feature detection and extracting based on Active Appearance Models (AAM) with Ekman's Facial Action Coding System (FACS). Our approach to facial emotion recognition lies in the dynamic and probabilistic framework based on Dynamic Bayesian Network (DBN). The active appearance model (AAM) is a well-known method that can represent a non-rigid object, such as face, facial expression. In this paper, our approach to facial feature extraction lies in the proposed feature extraction method based on combining AAM with Facial Action Units of Facial Action Coding System (FACS) for automatically modeling and extracting the facial emotional features. Also, we use the Dynamic Bayesian Networks (DBNs) for modeling and understanding the temporal phases of facial expressions in image sequences.

**Keywords:** Facial Emotion Recognition Facial Feature Extraction, Active Appearance Model, Facial Action Coding System, Dynamic Bayesian Network.

## 1 Introduction

Over the last decades, resolving the absence of mutual sympathy in interactions between humans and machines is one of the most important issues in human-computer interaction, owing to its wide range of applications in commerce. Amongst these issues, many researches about the emotional communication between the human and the machine have been reported. The emotional communication between the human and the machine have received large attention as parts of the human computer interactions.

Among the emotional factor, facial expression is one of the most important and spontaneous element to recognize the human emotional status. Facial expressions in spontaneous interactions are often characterized by the presences of significant head motion, temporal behaviors, partial occlusions. But, it is difficult to define the human's emotion states by using facial expression, because, these facial expression characteristics are sensitized to the external noises, such as illumination conditions, dynamical head motion, etc. To overcome these problems, numerous techniques have

---

\*This work was supported by Mid-career Researcher Program through NRF grant funded by the MEST (No. 2010-0027627).

been proposed, but it still failed to be improved. One of the fundamental issues about facial expression analysis is how to represent visual information, so that the system can reveal the subtle expression.

Most attempts on the representation of visual information for facial expression have focused on optical flow analysis from face action. Several researches have employed the model-based techniques, and recent years have seen the increasing use of feature geometrical analysis to represent visual facial information [1]. Also it is difficult to classify a defined emotion. Examples of emotions are very various as given by the experiments of Ekman as basic six emotions [2].

The preceded researches have shown that people categorize facial emotions in a similar way across cultures, that similar facial expressions tend to occur in response to particular emotion eliciting events, and that people can produce simulations of facial emotion that are characteristic of each specific emotion [3]. A significant amount of research on spatial analysis and classification for facial expression recognition has focused on using Neural Network, Gabor wavelet based method, LDA[4] [5].

So we need to develop an advanced ways to recognize the emotion by using a single channel such as through facial expressions. In this paper we propose an advanced facial emotional feature extraction method by using combining AAM with FACS and facial feature classification method based on Dynamic Bayesian Network.

## 2 Facial Emotional Feature Extraction

### 2.1 The Problem of Facial Emotion Feature Extraction

Precise localization of a facial feature plays an important role in feature extraction, and expression recognition. However in actual applications, because of the difference in a facial shape and the quality of the database image, it is difficult to locate the facial feature precisely [6]. In the face, we use the eyebrow and mouth corner shape as main ‘anchor’ points. There are many methods available for facial feature extraction, such as Eye Blinking Detection, Eye Location Detection, Segmentation of Face area and feature detection, etc [7]. In spite of this, facial feature extraction for recognition is still a challenging problem.

The problem of extracting the feature of interest including emotional information is analogous to the problem of image segmentation from certain image regions. There are two kinds of segmentation method: model-based segmentation and image-based segmentation. Image-based segmentation methods, such as the “Live-wire method”, are more accurate in feature boundary delineation, but they are poor in boundary recognition, such as feature region extraction by matching object areas in the total image [8].

In this paper, we accept the limitations associated with the degradation of feature delineation and use the one of the model-based segmentation methods, namely the Active Appearance Model (AAM).

### 2.2 Combining AAM with Facial Action Coding System

Active Appearance Model (AAM) is a statistical approach for shape and texture modeling and feature extraction. It has been extensively used in computer vision

applications to model the class-variability and motion of objects in images. It represents a target structure by a parameterized statistical shape and texture model obtained from training set. This method was introduced by Cootes et al. [9]. AAM is an extension of this approach rather than tracking a particular object, this models of appearance can match to many of a class of deformable objects (e.g., any face with any expression, rather than one persons face with a particular expression) [9].

AAM generates statistical appearance models by combining a model of shape variation with a texture variation. "Textures" are pixel intensities of the target image.

An object is described by points, referred to as landmark points. The landmark points are (manually) determined in a set of training images. From these collections of landmark points, a point distribution model [8] is constructed as follows. The landmark points are stacked in shape vectors

$$\mathbf{x} = [x_1, y_1, \dots, x_n, y_n]^T \tag{1}$$

Variability is modeled by means of a Principal Component Analysis (PCA) an eigenanalysis of the dispersions of shape and texture.  $\mathbf{s}$ ,  $\mathbf{t}$  denote a synthesized shape and texture and  $\bar{\mathbf{s}}$ ,  $\bar{\mathbf{t}}$  denote the corresponding sample means.

$$\mathbf{s} \approx \bar{\mathbf{s}} + \Phi \mathbf{b}_s, \quad \mathbf{t} \approx \bar{\mathbf{t}} + \Phi \mathbf{b}_t, \tag{2}$$

where here  $\mathbf{b}_s$ ,  $\mathbf{b}_t$  is an adjusting PC scores,

$$\bar{\mathbf{b}} = \begin{bmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{W}_s \Phi_s^T (\mathbf{s} - \bar{\mathbf{s}}) \\ \Phi_t^T (\mathbf{t} - \bar{\mathbf{t}}) \end{bmatrix} \tag{3}$$

When fitting the model to a set of points, the values of  $\mathbf{b}$  are constrained to lie within the range  $\pm m\sqrt{\lambda_i}$ , where  $m$  usually has a value between two and three. The number  $t$  of eigenvalues to retain is chosen so as to explain a certain proportion  $f_v$  of the variance in the training shapes, usually ranging from 90% to 99.5%. The desired number of modes is given by the smallest for which

$$\sum_{i=1}^t f_i \geq f_v \sum_{i=1}^{2n} \lambda_i \tag{4}$$

A suitable weighting between pixel distances and pixel intensities is carried out through the diagonal matrix  $\mathbf{W}_s$ . To recover any correlation between shape and texture the two eigenspaces are coupled through a third PC transform

$$\bar{\mathbf{b}} \approx \Phi_c \mathbf{c} \tag{5}$$

$$\mathbf{s} = \bar{\mathbf{s}} + \Phi_s \mathbf{W}_s^{-1} \Phi_{c,s} \mathbf{c}, \quad \mathbf{t} = \bar{\mathbf{t}} + \Phi_t \mathbf{W}_t^{-1} \Phi_{c,t} \mathbf{c} \tag{6}$$

Before PCA is applied to the shapes, the shapes can be aligned by translating, rotating and scaling them so as to minimize the sum of squared distances between the landmark points. An iterative scheme known as Procrustes analysis [9] is used to align the shapes. This alignment procedure makes the shape model independent of the size, position, and orientation of the objects. To regularize the model,  $\Phi_s, \Phi_t, \Phi_c$  are truncated [10].

AAM creates the shape, texture combination model of training facial image sequence, and by using the AAM fitting procedure, the appearance model can be adjusted to the input facial image. In order to analyze the feature region of the input shape model and encode facial feature movement, in this paper, we adopt the Facial Action Coding System (FACS). A FACS is a preferred basis for encoding facial feature movement and it developed by P. Ekman and W. V. Friesen is a detailed, technical guide that can explain how to categorize facial behaviors based on the muscles that produce them and how muscular action is related to facial appearances. Facial appearance changes are described in terms of 44 facial action units, each of which is related to the contraction of one or more facial muscles. The system illustrates appearance changes of the face using written descriptions, using still images as examples [11]. FACS measurement units are action units and, each Action Unit describes a facial muscle area that is related to an emotional region [11][12].

In this paper, we use an algorithm for emotional feature extraction and facial feature analysis by applying a combination of FACS and AAM techniques. Figure 1 shows the main action units regions including emotional feature information. Each region of facial shape model alphabetic indexed shows the eyebrow, eyelid, mouth lip regions for analyzing emotion feature of facial image.

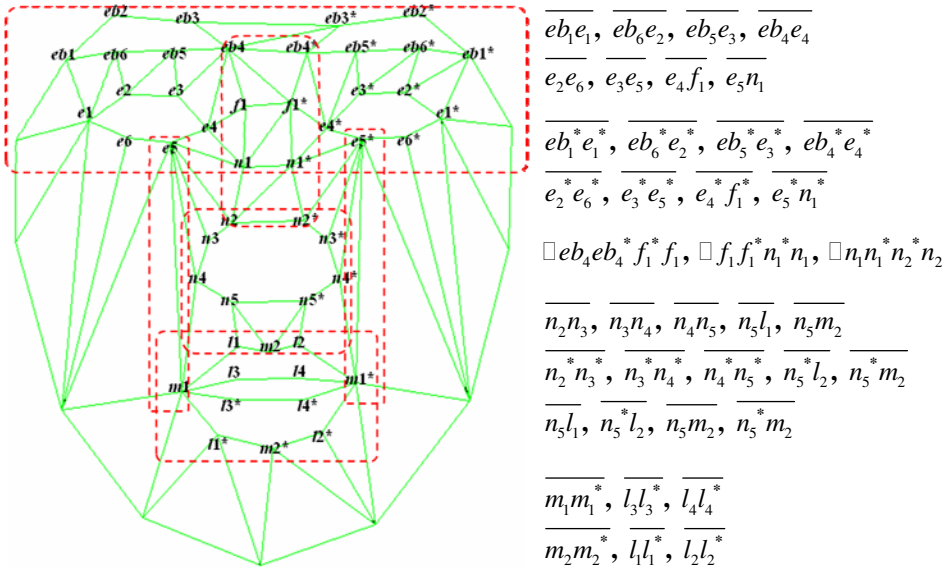


Fig. 1. Emotional Feature Parameters based on FAUs

### 3 Facial Emotion Recognition Based on DBN

We present a facial feature extraction method based on combining AAM with FACS. Also, we need to select a facial emotion recognition method. In this paper, we create a BN model to represent the causal relations between the facial expressions and

emotional features (facial AUs based on AAM) and extend the BN to a DBN model for modeling the dynamic behaviors of facial expressions in image sequences.

Each feature regions of a facial image sequence are chosen by the collection of FACS-action units, so that every face can be recognized even when the details of the individual features (eyelid, eyebrow, and mouth lip) are no longer resolved. In Figure 2, we use training image sequences that emotional feature regions are displayed by a set of yellow lines and landmark points.

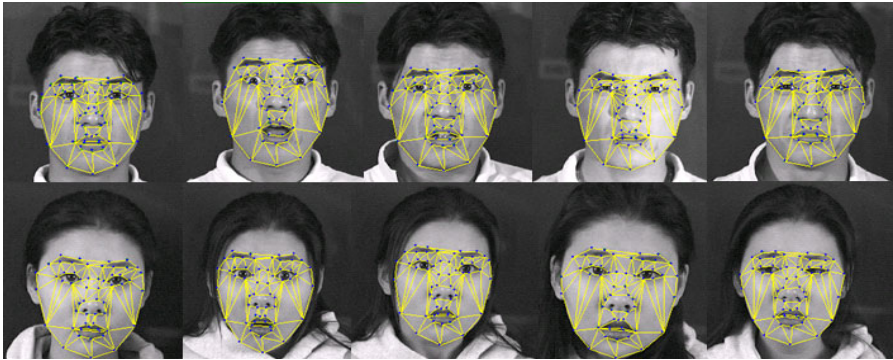


Fig. 2. Image Sequence Database for Emotion Recognition

And, following Figure 3 shows an example of a facial model obtained by proposed feature extraction method.

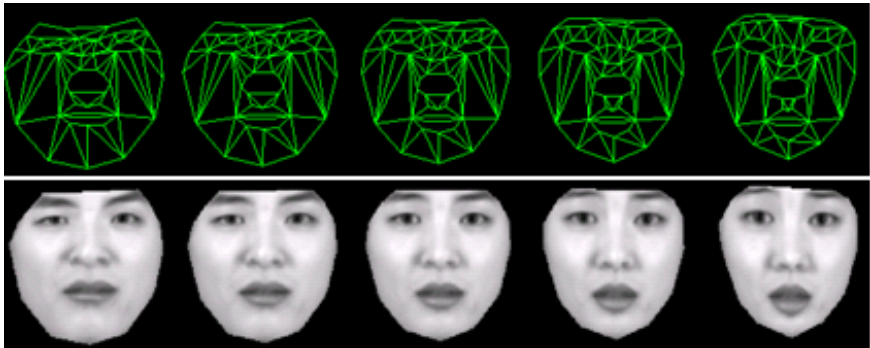
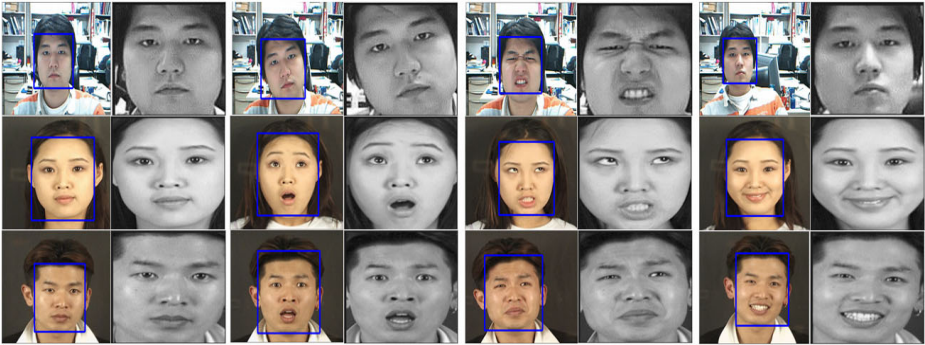


Fig. 3. Example of Active Appearance Model by using AAM

Numerous approaches have been proposed for facial feature tracking. In this paper, our approach to facial feature tracking relies on a haar-like feature model so that it can detect the outline of facial image under large variation in head orientation [13]. Figure 4 shows the result of haar-like feature detection of the image sequences.



**Fig. 4.** Haar-like facial feature detection

We present a classification method, which uses changes in facial emotion measurement values to detect action units and recognize six basic emotions in image sequences. Table 1 shows the FAUs and their interpretations. By using these descriptions, we can model a set of emotion feature parameters based on FAUs and facial visual cues shown in Figure 1.

**Table 1.** Facial Action Units and their Interpretations [2]
















<b>AU1</b>  Inner Brow raiser	<b>AU2</b>  Outer Brow raiser	<b>AU4</b>  Brow Lowerer	<b>AU5</b>  Upper lid raiser	<b>AU6</b>  Cheek raiser
<b>AU7</b>  Lid tightener	<b>AU9</b>  Nose wrinkler	<b>AU12</b>  Lip corner puller	<b>AU15</b>  Lip corner depressor	<b>AU17</b>  Chin raiser
<b>AU23</b>  Lip tightener	<b>AU24</b>  Lip presser	<b>AU25</b>  Lips part	<b>AU26</b>  Jaw drop	<b>AU27</b>  Mouth stretch

Figure 5 shows the proposed an emotion recognition procedure consists of two brief stages. Firstly, a facial feature should be extracted to get emotional information from the face region based on action units of FACS. Secondly the emotional states are classified using Dynamic Bayesian Network.

For the classification of the emotion from the appearance model of facial image sequences with extracting facial emotion feature information, primary, auxiliary action units are inserted to the static Bayesian Network.

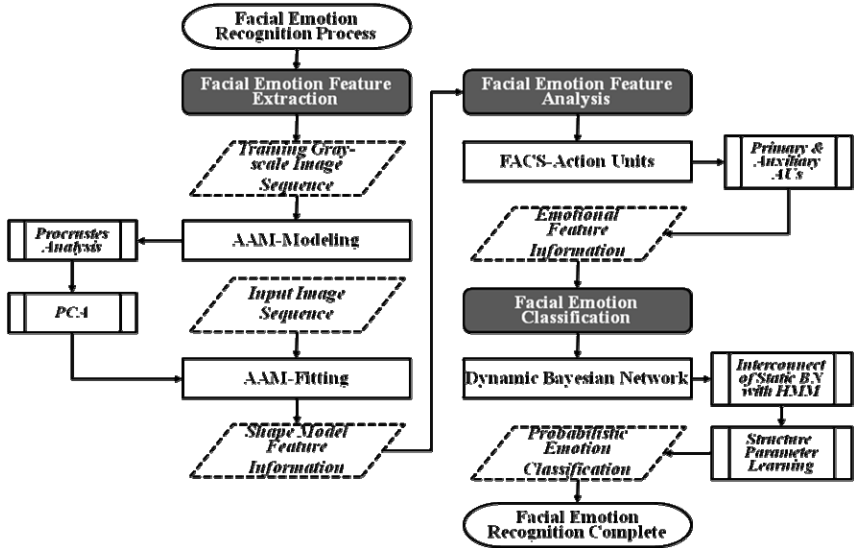


Fig. 5. Facial Emotion Recognition Algorithm

Following Figure 7 shows the Bayesian Network structure for emotion recognition. This structure consist of 3 node sets, and each nodes set are composed of primary, auxiliary FAUs, decision set of visual cues.

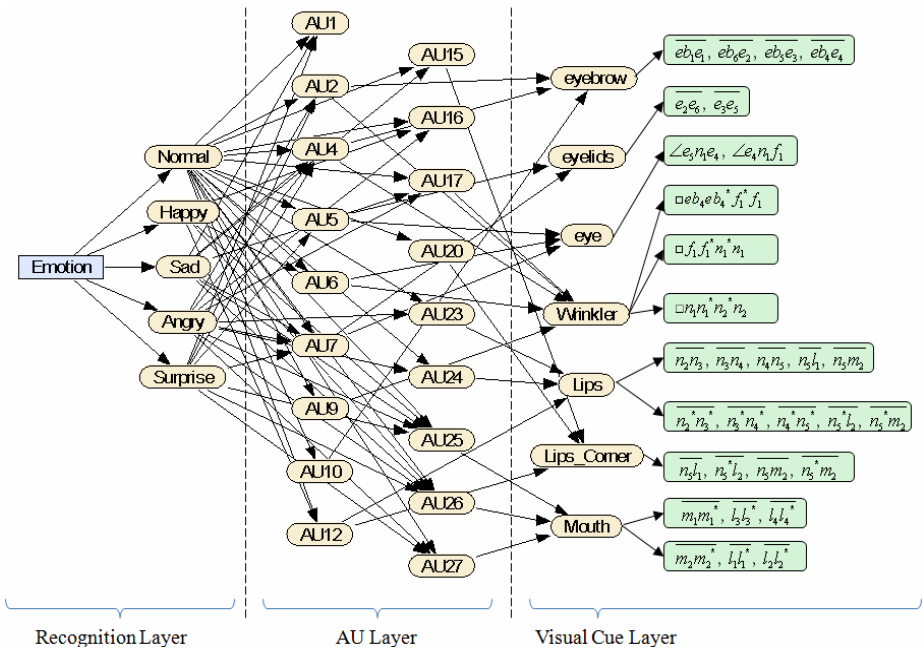


Fig. 6. Bayesian Network Structure for Emotion Recognition



Following Figure 8 shows the DBN based on temporal evolution of Bayesian Network structure.

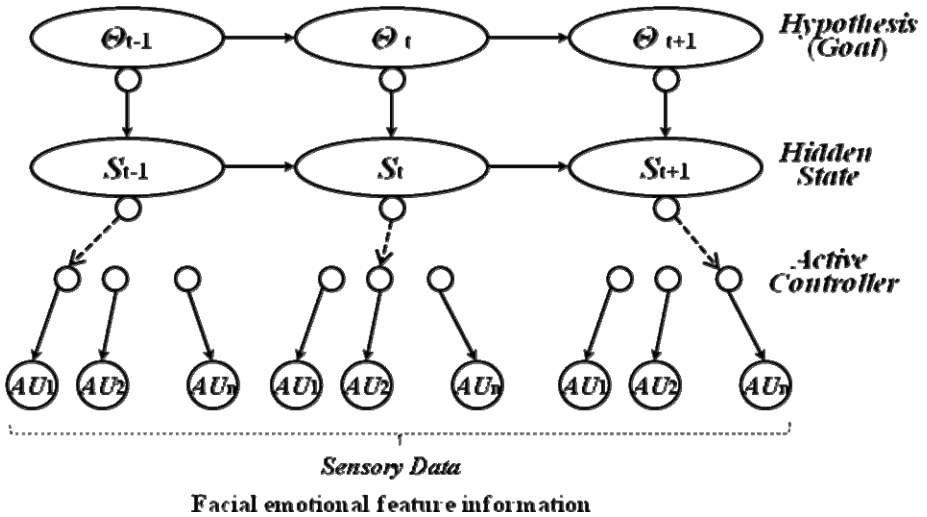


Fig. 7. Temporal link of DBN considering Temporal Evolution of Emotions

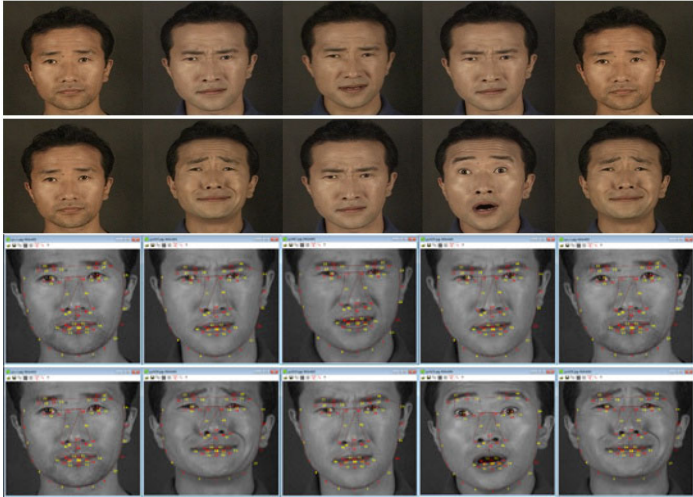
Following Figure 9 shows the emotion recognition simulation program. We have chosen the facial image database at DTU-AAM site (<http://www2.imm.dk/~aam>). In the simulation program, each picture control shows the result of emotional feature extraction, analysis, and recognition based on DBN with HMM



Fig. 8. Emotion Recognition Simulation Program

### 4 Experimental Result

This Section presents experimental results demonstrating the promise of the proposed approach. The following set of above experiments is used primarily for demonstrating performance characteristics of this approach. We create an image sequence involving multiple emotions shown Figure 10.



**Fig. 9.** A Segmentation of Emotional Feature of Facial Image Sequence

It can be seen that the temporal evolution of the emotion varies over frames, Figure 10 and Table 2 provide the analysis result by our facial emotion recognition approach. There is improving recognition performance compared with the based and holistic approach using optical flow fields.

**Table 2.** Classification Result(%) for the Image Sequence

Frame	Happiness	Sadness	Surprise	Anger	Normal
1	0.12	0.62	0.12	0.12	<b>0.85</b>
2	0.24	0.51	0.21	0.21	0.78
3	0.26	0.33	0.34	0.52	0.32
4	0.24	0.51	0.51	<b>0.96</b>	0.71
5	0.12	0.62	0.52	<b>0.75</b>	<b>0.96</b>
6	0.12	<b>0.61</b>	0.43	<b>0.74</b>	<b>0.95</b>
7	0.21	<b>0.91</b>	<b>0.95</b>	0.65	0.23
8	0.23	<b>0.61</b>	<b>0.94</b>	0.52	0.72
9	0.42	0.21	<b>0.72</b>	0.42	0.12
10	0.11	0.92	0.27	0.12	0.32

Table 2 represents the numerical results of probability distributions over image sequence frames. We can readily see from the result that the facial emotions in the rest of image frames are correctly classified. Also, we can see that the recognition performance is improving compared with the Table 3 method based on active sensing facial emotion feature extraction method and DBN [1].

**Table 3.** Classification Result(%) for the Image Sequence based on [1]

Frame	Happiness	Sadness	Surprise	Anger	Normal
1	0.167	0.167	0.167	0.167	<b>0.852</b>
2	0.005	0.371	0.086	0.521	0.485
3	0.003	0.364	0.071	0.786	0.121
4	0.005	0.460	0.086	<b>0.886</b>	0.112
5	0.067	0.467	0.067	<b>0.456</b>	<b>0.765</b>
6	0.107	<b>0.467</b>	0.007	<b>0.312</b>	<b>0.803</b>
7	0.061	<b>0.872</b>	<b>0.665</b>	0.025	0.231
8	0.005	<b>0.907</b>	<b>0.465</b>	0.207	0.429
9	0.002	0.304	<b>0.870</b>	0.002	0.128
10	0.061	0.712	0.665	0.035	0.320

## 5 Conclusion and Future Works

For facial emotion recognition, in this paper a facial emotional feature extraction procedure based on an AAM, FACS. The use of an AAM facilitates extraction of various emotional features region. Facial action units are represented by a set of geometric, appearance-based and surface-based measurements, which are effectively classified with respect to an emotional state. The method by combining the AAM with a FACS is robust about the initial state or with respect to the disturbance of database image sequences. Also, the DBN classifies the emotional states and it is expected that the proposed emotion recognition method performs robust classifications concerning the missing and uncertain data and temporal evolution of the image sequences.

It will be the subject of future work to apply advance feature extraction methods because the vulnerability of an AAM such as the necessity for a large number of landmarks and time data. Also, in order to classify emotions more delicately, it is required that the change in the structure of the Bayesian Network by using automatic structure learning and the conditional probability table of the Bayesian Network be considered in relation to structure learning and parameter learning. Furthermore, we need further study on the recognition rate improvement when the emotion feature regions are overlapping.

## References

1. Zhang, Y., Ji, Q.: Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(5), 699–714 (2005)
2. Ekman, P., Friesen, W.V.: The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* 1, 49–98
3. DataFace Site: Facial Expressions, Emotions Expressions, Nonverbal Communication, Physiognomy, <http://www.face-and-emotion.com>
4. Padgett, C., Cottrell, G.: Representing Face Images for Emotion Classification. In: Mozer, M., Jordan, M., Petsche, T. (eds.) *Advances in Neural Information Processing Systems*, vol. 9 (1997)
5. Lyons, M.J., Budynek, J., Akamatsu, S.: Automatic Classification of Single Facial Images. *IEEE Trans. Pattern Analysis and Machine Intelligence* 21(12), 1357–1362 (1999)
6. Yi-Bin, S., Jian-Ming, Z., Jian-Hua, T., Geng-Tao, Z.: An improved facial feature localization method based on ASM. In: *7th International Conference on Computer-Aided Industrial Design and Conceptual Design, CAIDCD 2006* (2006)
7. Kobayasho, S., Hashimoto, S.: Automated feature extraction of face image and its applications. In: *International Workshop on Robot and Human Communication*, pp. 164–169
8. Liu, J., Udupa, J.K.: Oriented Active Shape Models. *IEEE Transactions on Medical Imaging* 28(4), 571–584 (2009)
9. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(6), 681–685 (2001)
10. Stegmann, M.B.: *Analysis and Segmentation of Face Images using Point Annotations and Linear Subspace Techniques*. informatics and mathematical modelling Lingby, Denmark: Tech. Univ. Denmark, IMM Tech. Rep. IMM-REP-2002-22 (August 2002)
11. Ekman, Friesen, W.: *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto (1978)
12. Ekman, P., Friesen, W.V.:  
<http://www.cs.cmu.edu/afs/cs/project/face/www/facs.htm>
13. Mita, T., Kaneko, T., Hori, O.: Joint Haar-like Features for Face Detection. In: *Proceedings of the 10th IEEE International Conference on Computer Vision*, October 17–21, vol. 2, pp. 1619–16261 (2005)

# An Analysis of Node Localization Error Behavior in Wireless Sensor Networks

Dan Hai, Hui Zhang, and Zhiqiang Zheng

Department of Automatic Control, National University of Defense Technology  
410073 Changsha, Hunan, China  
haidan\_hd@hotmail.com, zhanghui\_nudt@126.com,  
zqzheng@nudt.edu.cn

**Abstract.** Localizing the sensor nodes is a fundamental problem in wireless sensor networks (WSN). In this paper, we explore the localization error behavior with respect to the network setup and measurement noise in case of using range measurement to localize the network. Our analysis is based on analytical bounds for the covariance given by the CRB. We expand the classical application CRB to a formulation which allows us to compute the uncertainty associated with the node for the case of beacon with uncertain positions, and then develop a fundamental understanding of how network setup parameters and measurement noise affect the error behavior in node position estimates. The goal of this paper is to quantify the network setup parameters and measurements for minimizing the node location error in WSN application.

**Keywords:** Sensor Node Localization, Cramé Rao Bound, Uncertain Beacons.

## 1 Introduction

The recent growth in wireless communication and sensing technologies has emphasized the importance and applicability of Wireless Sensor Networks (WSN) for a wide variety of application domains. In general, sensor networks are mostly used for monitoring purposes, so sensor's location has to be accurately determined. Sensing data without knowing the sensor's location is meaningless [1]. On the other hand, if sensor nodes were well located, such a sensing system could be very useful to provide services such as sensing and tracking target efficiently. Therefore, the localization of the sensor nodes for sensor network deployment is an essential problem for WSN.

Getting the node positions from some location devices such as GPS is a good way, but it is not available for every node due to the cost, energy consumption and the satellite invisibility. Various methods have been proposed to solve the localization problem in WSN. In general most methods are based on a small and well distributed set of nodes with known positions, called beacon. Nodes with unknown locations use advertised beacon locations and a set of distance or angular measurements to their neighboring nodes to estimate their location. The localization algorithm is based on triangulation, maximum bounding or other techniques [2,3,4].

Even in idealized setup with no obstacle or other external factors, relatively small error from noisy measurements can induce much larger errors in node position estimates. So the quality of location estimates has received considerable attention in the literature.

The Cramé Rao Bound (CRB) [5] gives a lower bound on the error covariance matrix for an unbiased estimate of parameter vectors. It is widely used to evaluate the accuracy of any unbiased estimators and provides a performance benchmark. CRB allows determination if accuracies necessary for particular application are possible. It has already been derived in [6].

Most early work and applications using CRB assumed the position of beacons to be error free. In case error associated with beacon location is significantly smaller than the error associated with the measurement, it could therefore be ignored. Otherwise the error from beacon location may induce much larger errors in node position estimates. In real application, the beacon location was obtained by initial calibration. For some affection from environment or intrinsic error in data obtained by calibration sensor, the prior location of beacon may contain error can not be ignored. So we should explore how does beacon uncertainty affects the node position estimates.

In this paper, we focus on the possible accuracy using any unbiased relative location estimator. Our study is interest in the localization error behavior with respect to the network setup in case of using range measurement to localize the network.

Our analysis is based on analytical bounds for the covariance given by the CRB. We extend the application of CRB for the case in which there is beacon uncertainty. Motivated by error model for robot pose estimation based on triangulation in Multi-robot localization problem [7,8], we provide a fundamental understanding of how the beacon's distribution, the position uncertainty of the beacon and the measurement noise affects the magnitude of node location estimate uncertainty in system by using CRB. The goal of this paper is to quantify the network setup parameters and measurements for minimizing the node location error in WSN application.

## 2 Problem Statement

We assume a set of  $n$  sensor nodes in a  $x$ - $y$  plane, with unknown locations. In this paper, we called the node whose location is unknown and to be calculated as "localizing-node". In addition, a set of  $m$  sensor nodes spread in the plane as beacons, with a prior information about their locations but the locations have some uncertainty. We call them as "referenced-node" to make a distinction to beacon whose location is determined.

We define the location parameter vector  $\theta = [\theta_N \ \theta_M]^T$ . Where  $\theta_N$  contains localizing-node location parameters which is unknown,  $\theta_M$  contains referenced-node location parameters.  $\theta_N = [\theta_1, \dots, \theta_n]^T$ ,  $\theta_M = [\theta_{n+1}, \dots, \theta_{n+m}]^T$ , where the parameter  $\theta_i = (x_i, y_i)$ .

For node localization, each node advertises its location to the other nodes neighboring in the network and emits a known signal that allows neighboring nodes to estimate their distance to the emitting node. We denote neighbor set of node  $i$  is

$E(i)$ , define a set  $L = \{(i, j) | i \in N, j \in E(i)\}$  which is composed of pairs of nodes that are within the measurement range. We use the indicator function  $I(i, j)$ , where  $I(i, j) = 1$  indicates that a distance measurement are observed between node  $i$  and node  $j$  and  $I(i, j) = 0$  on the contrary. The distance measurement at node  $i$  to node  $j$  is defined as  $\bar{d}_{i,j}$ , and distance measurements set of system is defined as  $\bar{d}$ ,  $\bar{d} = \{\bar{d}_{i,j} | (i, j) \in L\}$ .

$$\bar{d}_{i,j} = d_{i,j} + e_{i,j} \tag{1}$$

$$d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{2}$$

Where  $e_{i,j}$  is the distance measurement error and  $d_{i,j}$  is the true distance between nodes  $i$  and  $j$ . We assume the measurements are independent to each other and  $\bar{d}_{i,j} = \bar{d}_{j,i}$  by reciprocity.

The probability density function of distance measures can be described as:

$$p(\bar{d} | \theta) = \prod_{(i,j) \in L} p(\bar{d}_{i,j} | \theta_i, \theta_j) \tag{3}$$

We define  $p \triangleq p(\bar{d} | \theta)$ ,  $p_{i,j} \triangleq p(\bar{d}_{i,j} | \theta_i, \theta_j)$ , so the log form of Equation(3) can be described as follows in brief:

$$\ln p(\bar{d} | \theta) = \ln \left[ \prod_{(i,j) \in L} p(\bar{d}_{i,j} | \theta_i, \theta_j) \right] = \ln p = \sum_{(i,j) \in L} \ln p_{i,j} \tag{4}$$

The CRB is used to bound localization error. When beacon nodes have known location, the standard CRB is defined as  $Cov(\theta_N) \geq CRB(\theta_N)$ , where  $Cov(\theta_N)$  is the error bound of unknown parameters estimate,  $Cov(\theta_N) = E\{(\hat{\theta}_N - \theta_N)(\hat{\theta}_N - \theta_N)^T\}$ ,  $\hat{\theta}_N$  is the estimate of  $\theta_N$ .

The lower bound is given in terms of Fisher Information Matrix (FIM) [5]:  $F(\theta_N)$ , where  $CRB(\theta_N) = [F(\theta_N)]^{-1}$  and  $F(\theta_N) = -E\{\nabla_{\theta_N} [\nabla_{\theta_N} (\ln p)]^T\}$ .

### 3 Node Localization with Uncertain Beacon Locations

In case the referenced-node locations have some uncertainty, we assume the uncertainty can be described as a prior pdf  $p_0(\hat{\theta}_M | \theta_M)$ . This prior knowledge provides useful information about referenced-node location, it can be thought as a measurement from “pseudo-beacon” whose location is determined to localize referenced-nodes. Based on this assumption, we model the referenced-node location parameter vector  $\theta_M$  as a random vector, and expand the CRB to the full state parameter vector  $\theta$ .

We denote  $p_0 \triangleq p_0(\hat{\theta}_M | \theta_M)$  in brief. For the independence between  $p$  and  $p_0$ , CRB is given as:

$$CRB(\theta) = [F(\theta)]^{-1} \tag{5}$$

$$F(\theta) = F(\theta_N, \theta_M) = -E\{\nabla_\theta[\nabla_\theta \ln(p \bullet p_0)]^T\} \tag{6}$$

Equation (6) can be divided:  $F(\theta) = -E\{\nabla_\theta[\nabla_\theta \ln p]^T\} - E\{\nabla_\theta[\nabla_\theta \ln p_0]^T\}$ .

We denote  $F_N(\theta) = -E\{\nabla_\theta[\nabla_\theta \ln p]^T\}$ ,  $F_0(\theta) = -E\{\nabla_\theta[\nabla_\theta \ln p_0]^T\}$ ,

$$F(\theta) = \begin{bmatrix} f_{1,1} & \cdots & f_{1,n+m} \\ \vdots & \ddots & \vdots \\ f_{n+m,1} & \cdots & f_{n+m,n+m} \end{bmatrix}, \quad F_N(\theta) = \begin{bmatrix} q_{1,1} & \cdots & q_{1,n+m} \\ \vdots & \ddots & \vdots \\ q_{n+m,1} & \cdots & q_{n+m,n+m} \end{bmatrix},$$

$$F_0(\theta) = \begin{bmatrix} l_{1,1} & \cdots & l_{1,n+m} \\ \vdots & \ddots & \vdots \\ l_{n+m,1} & \cdots & l_{n+m,n+m} \end{bmatrix} \tag{7}$$

Where  $f_{i,j} = q_{i,j} + l_{i,j}$

$$q_{i,j} \text{ was given as } q_{i,j} = \begin{cases} -\sum_{(i,k) \in L} I(i,k) \bullet E[\partial^2 p_{i,k} / \partial \theta_i^2] & (i = j) \\ -I(i,j) \bullet E[\partial^2 p_{i,j} / \partial \theta_i \partial \theta_j] & (i \neq j) \end{cases}$$

Partition Equation (7) into blocks, we get  $\begin{bmatrix} G_{n \times n} & U_{n \times m} \\ V_{m \times n} & S_{m \times m} \end{bmatrix}$ . Because  $p_0$  has no relationship with localizing-node,  $U_{n \times m} = (V_{m \times n})^T = 0$ ,  $G_{n \times n} = 0$ ,

$$S_{m \times m} = F(\theta_M) = -E\{\nabla_{\theta_M}[\nabla_{\theta_M} \ln p_0]^T\} \tag{8}$$

### 3.1 Gaussian Assumption

To simplify the analysis of error propagation problem, we assume error associated with the distance measurement as a Gaussian white noise and referenced-node  $i$  prior distribution is bivariate Gaussian distribution. Even the prior distribution model of referenced-node can apply to more general case such that some referenced-nodes location error are correlated, we assume the error of referenced-node is independent in this paper. The Gaussian assumption is not always a valid model for the error distribution, but it can be used to provide a conservative estimate if the true error distribution is similar to a Gaussian.

We can obtain the following conditional probability density function as :

$$p_{i,j} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(\bar{d}_{i,j} - d_{i,j})^2}{2\sigma^2}\right\} \tag{9}$$

$$p_0(\hat{\theta}_i | \theta_i) = \frac{1}{\sqrt{2\pi} |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}[\hat{\theta}_i - \theta_i]^T \Sigma_i^{-1} [\hat{\theta}_i - \theta_i]\right\} \tag{10}$$



with covariance matrix  $\Sigma_i = \begin{bmatrix} \sigma_{i,xx}^2 & \sigma_{i,xy}^2 \\ \sigma_{i,yx}^2 & \sigma_{i,yy}^2 \end{bmatrix}$ .

From the description in previous section, we got:

$$l_{i,j} = \begin{cases} \Sigma_i^{-1} & (i = j) \\ 0 & (i \neq j) \end{cases}, q_{i,j} = \begin{bmatrix} q_{(i,j),xx} & q_{(i,j),xy} \\ q_{(i,j),yx} & q_{(i,j),yy} \end{bmatrix}, \text{ where}$$

$$q_{(i,j),xx} = \begin{cases} -\sum_{(i,k) \in L} I(i,k) \cdot E \left[ \partial^2 p_{i,k} / \partial x_i^2 \right] & (i = j) \\ -I(i,j) \cdot E \left[ \partial^2 p_{i,j} / \partial x_i \partial x_j \right] & (i \neq j) \end{cases}$$

$$q_{(i,j),yy} = \begin{cases} -\sum_{(i,k) \in L} I(i,k) \cdot E \left[ \partial^2 p_{i,k} / \partial y_i^2 \right] & (i = j) \\ -I(i,j) \cdot E \left[ \partial^2 p_{i,j} / \partial y_i \partial y_j \right] & (i \neq j) \end{cases}$$

$$q_{(i,j),xy} = q_{(i,j),yx} = \begin{cases} -\sum_{(i,k) \in L} I(i,k) \cdot E \left[ \partial^2 p_{i,k} / \partial x_i \partial y_i \right] & (i = j) \\ -I(i,j) \cdot E \left[ \partial^2 p_{i,j} / \partial x_i \partial y_j \right] & (i \neq j) \end{cases}$$

### 3.2 Metric

We used two metrics as tools for classifying and determining this model’s performance. The metrics are:

1) The covariance ellipse. The ellipse is related to the covariance matrix  $C(\theta_i)$  and whose semi-minor and semi-major axis lengths are two standard deviations. The area of the ellipse is 95% confidence of the estimate.  $C(\theta_i)$  is defined as follow:

$$C(\theta_i) = [F(\theta)]_{i,i}^{-1} \tag{11}$$

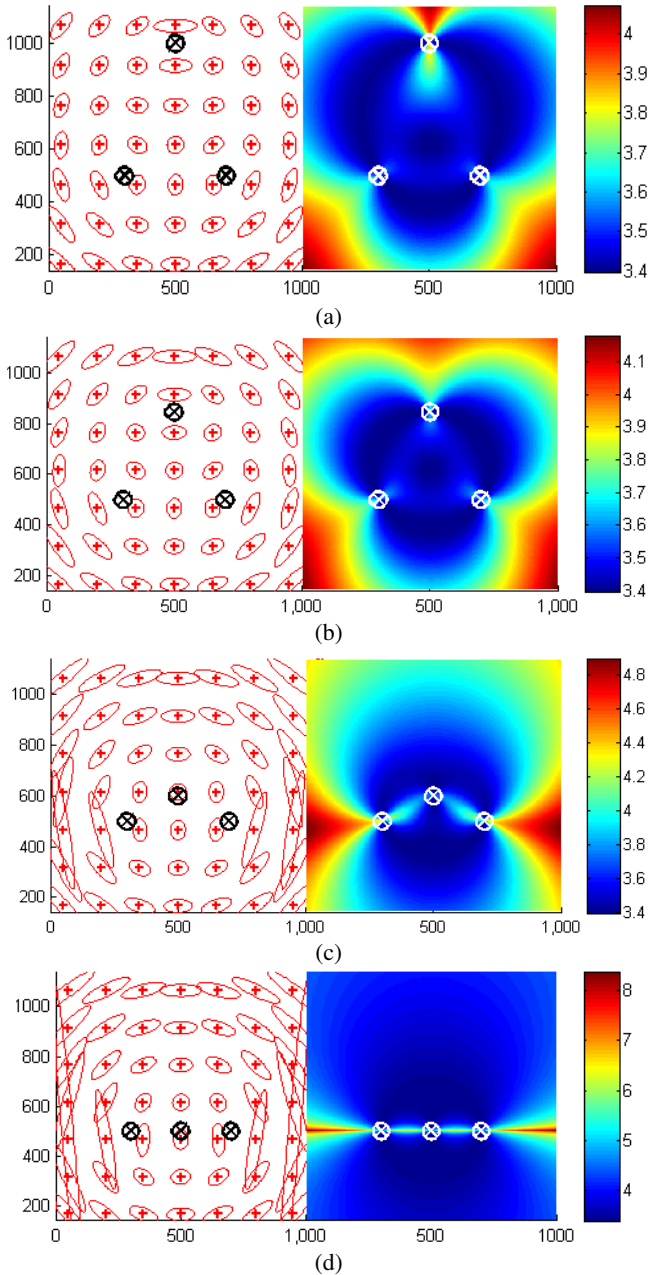
2) The trace of the covariance of the  $i$  th location estimate  $tr\{C(\theta_i)\}$ . It is a single value that can be used to quantify the overall error of bivariate Gaussian distribution about the node. It does not provide any information about the shape of the error. It satisfies

$$tr\{C(\theta_i)\} = tr\{[F(\theta)]_{i,i}^{-1}\} = [F(\theta)]_{(i,i),xx}^{-1} + [F(\theta)]_{(i,i),yy}^{-1} \tag{12}$$

## 4 Experimental Analysis

In this section, we will testify the covariance bound derived in the previous section. Several scenarios were designed to explore covariance bound behaves under different conditions as well as to understand the results of it.

We choose the following setup: there are 3 referenced-nodes and 1 localizing-node in the network. We set up a 1000 by 1000 point grid in the scenarios. The localizing-node placed at all grid points  $X = (i, j)$  in 10 unit-increments. We then compute the metrics at  $X = (i, j)$ .



**Fig. 1.** Metrics associated with localization of localizing-node (red plus in left) for a given set of three referenced-nodes and their associated position uncertainty (black in left and white in right). The left is metric 1 (red ellipse) defined in Equation (11), and the right is metric 2 (color map) in Equation (12) represented by log form. The possible position of localizing-node is red plus in left and for every  $1 \times 1$  grid of the plane in right.

### 4.1 Geometry Affection

Firstly, we will analyze the localization properties due to the network geometry configuration. As showed from Fig.1(a) to Fig.1(d), referenced-nodes 1 and 2 are fixed with  $\theta_{M1}=[300 \ 500]^T$ ,  $\theta_{M2}=[700 \ 500]^T$ . Then referenced-node 3 was moved from point a:  $\theta_{M3}=[500 \ 600+200\sqrt{3}]^T$  to b:  $\theta_{M3}=[500 \ 500+200\sqrt{3}]^T$ , c:  $\theta_{M3}=[500 \ 600]^T$  and d:  $\theta_{M3}=[500 \ 500]^T$ . The measurement noise and referenced-nodes prior uncertainty are fixed as  $\sigma^2 = 8^2$  and

$$\Sigma_{M1} = \Sigma_{M2} = \Sigma_{M3} = \begin{bmatrix} 12^2 & 0 \\ 0 & 12^2 \end{bmatrix}^T$$

localizing-node across and down the x-y plane at steps of unit-increments.

The color maps in Fig.1 show the metrics correspondingly. From the metrics show in those figures, we can see some universal characteristics: The area enclosed by the three referenced-nodes forms a plateau of minimum error with an additional lobe forming on the outside of the triangle formed by the referenced nodes. After a small initial distance within the area enclosed by the referenced, the rate of error increase rises sharply as the localizing-node moves further outside the area.

We can also see that the equilateral configuration seems to provide symmetrical result. In a way, the accuracy decreasing as the configurations get further away from the equilateral case. We can see that accuracy decreasing significantly with beacons to be a line arrangement. For the semi-major axis according to metric is increasing sharply.

### 4.2 Measurement Noise

We now further examine the metrics of the model for different measurement noise but same referenced-nodes configurations. We initially configure three referenced-node in an equilateral triangle arrangements with locations  $\theta_{M1}=[300 \ 500]^T$ ,

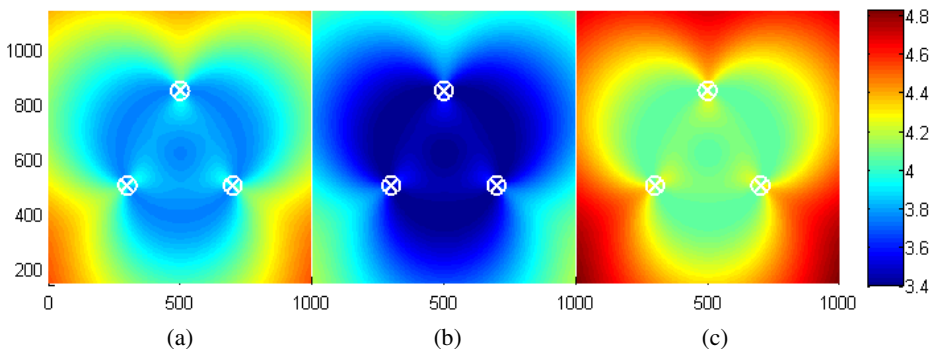


Fig. 2. Metric 2 comparison of different measurement noise affection in same scale

$\theta_{M_2} = [700 \ 500]^T$ ,  $\theta_{M_3} = [500 \ 500 + 200\sqrt{3}]^T$ , referenced-nodes prior uncertainty is

$$\Sigma_{M_1} = \Sigma_{M_2} = \Sigma_{M_3} = \begin{bmatrix} 12^2 & 0 \\ 0 & 12^2 \end{bmatrix}^T. \text{ The measurement noise is set to be a: } \sigma_a^2 = 18^2,$$

b:  $\sigma_b^2 = 8^2$  and c:  $\sigma_c^2 = 28^2$ .

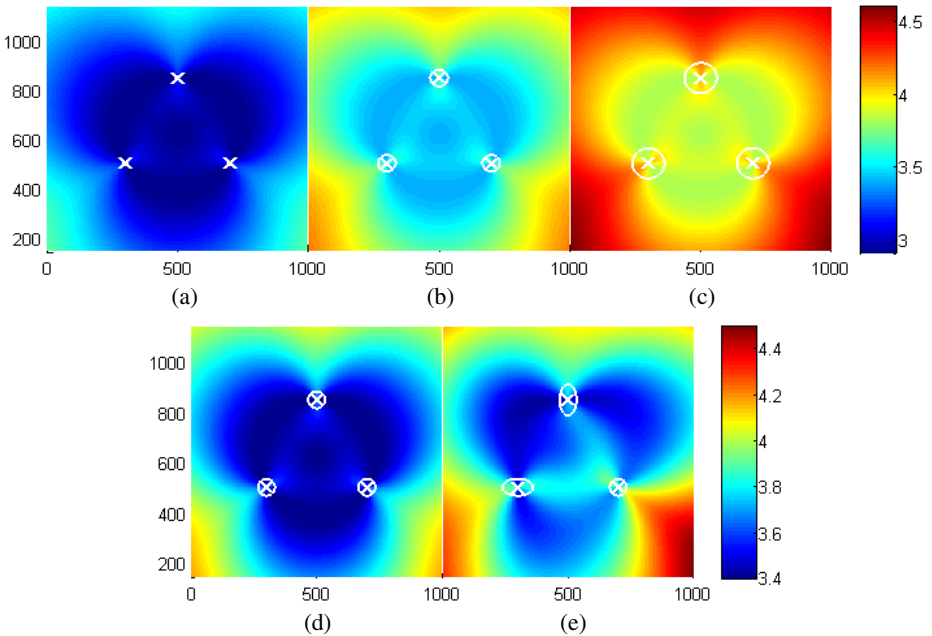
The metric 2 of localizing-node was compared at same scales in Fig.2, it produce the effect that the measurement noise perform among the cases mentioned above. From the figure, we will see that as measurement noise increasing the localization error of localizing-node increases at all points of the plane.

### 4.3 Referenced-Node Uncertainty

Knowing the uncertainty associated with the position of referenced-node, we can now quantify how it affects the uncertainty of the localizing-node position. In the experiments as showed in Fig.3(a)~(e) measurement noise is fixed to  $\sigma^2 = 8^2$ , the refer-

enced-nodes uncertainty is set to be a:  $\Sigma_{M_1} = \Sigma_{M_2} = \Sigma_{M_3} = \begin{bmatrix} 2^2 & 0 \\ 0 & 2^2 \end{bmatrix}^T$ , b:

$$\Sigma_{M_1} = \Sigma_{M_2} = \Sigma_{M_3} = \begin{bmatrix} 12^2 & 0 \\ 0 & 12^2 \end{bmatrix}^T, \text{ c: } \Sigma_{M_1} = \Sigma_{M_2} = \Sigma_{M_3} = \begin{bmatrix} 22^2 & 0 \\ 0 & 22^2 \end{bmatrix}^T.$$



**Fig. 3.** Metric 2 comparison of different referenced-node uncertainty affection in same scale

As soon as we change the error magnitude of referenced nodes, we can see that the areas of lower accuracy increase significantly with the error. What interests us is the metric change just like the case in section 4.2. So it is proved that in case error associated with beacon location has equal scale as the error associated with the measurement, beacon error could not be ignored. Otherwise it may induce larger errors in node position estimates.

We now further examine the affection of different referenced-node error to the metrics of the localization accuracy. Three referenced-nodes are in an

equilateral triangle arrangement. The prior error is  $e: \Sigma_{M1} = \begin{bmatrix} 22 \times 22 & 0 \\ 0 & 12 \times 12 \end{bmatrix}^T$ ,

$\Sigma_{M2} = \begin{bmatrix} 12 \times 12 & 0 \\ 0 & 12 \times 12 \end{bmatrix}^T$ ,  $\Sigma_{M3} = \begin{bmatrix} 12 \times 12 & 0 \\ 0 & 22 \times 22 \end{bmatrix}^T$ . The result is compared with d:

$\Sigma_{M1} = \Sigma_{M2} = \Sigma_{M3} = \begin{bmatrix} 12^2 & 0 \\ 0 & 12^2 \end{bmatrix}^T$  at same scale in Fig.3(d)-(e). We show the distribu-

tion of the referenced-node uncertainty not only affects the magnitude of localizing-node position uncertainty, but also the localizing-node position uncertainty distribution.

#### 4.4 Comparison with Classical Methods

In a sense, our method is the extension of the method in [6,7,8]. So it should provide the same result as the method described in the articles above under the same

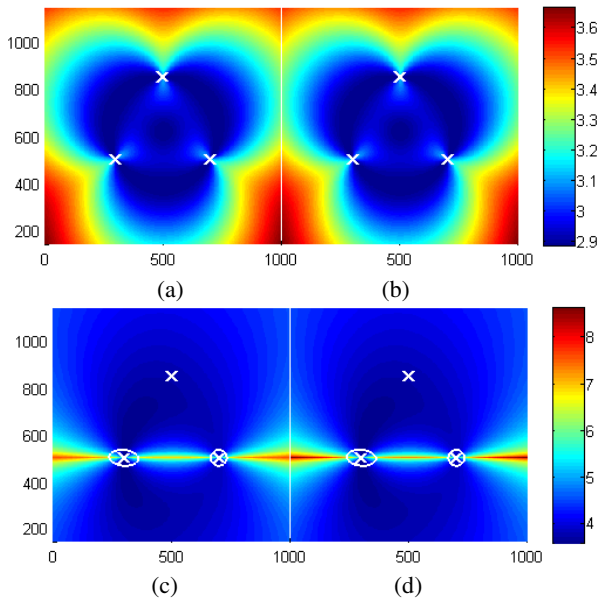


Fig. 4. Metric 2 comparison with classic methods

configuration of Network. In light of this, we propose to evaluate the model's performance and then present two sets of experiments with three beacons configuration.

One of the experiments is set measurement noise similar to section 4.3. We are using the method proposed in [6] and assumed the position of beacons to be error free. Correspondingly we set the prior error to be significantly small for approximating to error free in our method. The result shown in Fig.4(a)~(b) depicts the same effect of the two methods.

The second experiment is set two beacons localization scenario in trilateration in [8]. The uncertainty of referenced-node 1  $\Sigma_{M1}$  with  $\sigma_{M1,xx}^2 = 22^2$  and  $\sigma_{M1,yy}^2 = 12^2$  which is rotated by  $85^\circ$ . The uncertainty of referenced-node 2  $\Sigma_{M2}$  with  $\sigma_{M2,xx}^2 = \sigma_{M2,yy}^2 = 12^2$ . In our method, we set the third referenced-node uncertainty to be significantly large for approximating to absent of third beacon. The result was shown in Fig.4(c)~(d). It depict that the almost same effect happened using the two methods.

So the experiments provide the validity of our method, it proved that our method is the extension of the method in [6,7,8].

## 5 Conclusion and Future Work

In this paper we have derived and presented a CRB based model which can be used to evaluate the node localization estimate in WSN. We expand the CRB to a formulation which allows us to compute the uncertainty associated with the node for the case of beacon with uncertain positions. It is benefit to analyze error model in system. We proved that the accuracy of the node location estimate depends on beacon geometry, beacon location uncertainty and sensor noise.

We are currently in the process of extending this paper's research to the localization problem in the system composed by robots and WSN. We are particularly interested in the case where robots and WSN cooperate localizing each other. Overall, we believe that the model presented in this work is an effective tool which can be used in conjunction with localization techniques to deal with the problems such as robot path planning or node deployment.

## References

1. Rabaey, J.M., Ammer, M.J., da Silva, J.L., Patel, D., Roundy, S.: PicoRadio Supports Ad Hoc Ultra-low Power Wireless Networking. *IEEE Computer Magazine* 33(7), 42–48 (2000)
2. Chu, H.C., Jan, R.H.: GPS-less low-cost outdoor localization for very small devices. *IEEE Personal Communications* 7(5), 28–34 (2000)
3. Terwilliger, M., Gupta, A., Bhuse, V., Kamal, Z.H., Salahuddin, M.A.: A Localization System Using Wireless Network Sensors: A comparison of two techniques. In: *WPNC 2004*, Hannover, Germany (2004)
4. Li, Z., Trappe, W., Zhang, Y., Badri, N.: Robust Statistical Methods for Securing Wireless Localization in Sensor Networks. In: *IPSN 2005*, pp. 91–98 (2005)

5. Patwari, N., Ash, J.N., Kyperountas, S., Hero, A.O., Moses, R.L., Correal, N.S.: Location the Nodes: Cooperative Localization in Wireless Sensor Networks. *IEEE Signal Proc. Magazine*, 54–69 (2006)
6. Patwari, N., Hero, A.O., Perkins, M., Correal, N.S., O’Dea, R.J.: Relative Location Estimation in Wireless Sensor Networks. *IEEE Trans. Signal Processing* (2003)
7. Bahr, A., John, J.: Leonard: Minimizing Trilateration Error in the Presence of Uncertain Landmark Positions
8. Easton, A., Cameron, S.: A Gaussian Error Model for Triangulation-based pose estimation Using Noisy Landmarks. In: *Proc. IEEE International Conference on Robotics, Automation and Mechatronics*, Bangkok, Thailand (2006)

# Consensus Target Tracking in Multi-robot Systems

Zongyao Wang<sup>1</sup>, Dongbing Gu<sup>2</sup>, Tao Meng<sup>1</sup>, and Yanzhi Zhao<sup>1</sup>

<sup>1</sup> Surrey International Institute, Dongbei University of Finance and Economics,  
Dalian, China

qianxilong@googlemail.com

<sup>2</sup> School of Computer Science and Electronic Engineering, University of Essex,  
Colchester, UK  
dgu@essex.ac.uk

**Abstract.** By sharing collective sensor information, individuals in biological flocking systems have more opportunities for finding food and avoiding predators. This paper introduces a distributed robot flocking system with similar behaviour to biological flocking systems. In the developed flocking system, robots cooperatively track a target by using consensus algorithm. The consensus algorithm enables the robots to estimate locally the position of a target. The performance of the flocking system is tested via simulations. The results demonstrate that the flocking system is flexible, reliable and feasible for practical uses.

## 1 Introduction

In 1986, C. Reynolds created the first flocking model for computer simulation which is called “boids” [1]. Inspired by “boids” behaviours, multi-robot systems have become an active research area in the past few years. A migrating flocking system led by an experienced leader is a common phenomenon in biological colonies. As a consequence of evolution, the leader-follower strategy of the biological flocking systems has become an almost perfect mechanism. Like the biological flocking systems, a robot flocking system should be capable of moving along a predetermined route and reaching the specific destination as well. This task can be achieved by driving the robots in the flocking system to track a target [2]. In this paper, the leaders in a flocking system are the robots who can get the information of target’s state. The others are followers.

The challenging problem of target tracking is the consensus about the estimate of the target position among robots. Because of imprecise measurement, the robots could get different estimates about the target position. Due to the disagreement, it makes the flocking control more difficult. The flocking system could easily collapse because robots track “different targets”. Thus a cooperative estimation of the target position is necessary. We use a Kalman filter to implement the target estimation. The agreement is achieved via a consensus algorithm embedded in the Kalman filter. Each robot in the flocking system



can benefit from the observation of other robots and accordingly increase the tracking performance.

The Kalman filter is frequently used to estimate the state of a dynamic system from a series of noisy measurements. It is named after Rudolf Kalman who first introduced the Kalman filter in [3] (1960). The earliest application of the Kalman filter is to estimate the trajectory for the “Apollo program” by NASA Ames Research. Nowadays, the Kalman filter has been widely used in control systems and electronic systems.

In recent years, the consensus algorithm based Kalman filter has been investigated widely. To implement the data fusion in sensor networks, the consensus algorithm is used in [4]. By using wireless communication with neighbours, each sensor node obtains a measurement result from local neighbours and estimates a global result by using the consensus algorithm. Simulations are provided to verify that the sensor network can reach an agreement about the measurement. The consensus algorithm is also used to synchronise the coupled oscillators networks in paper [5]. It is proved in the paper that the agents can reach agreement about the frequency of oscillation. In [6], a consensus Kalman filter with the purpose of optimising two design parameters (the Kalman gain and the consensus matrix) is proposed.

In some cases, the sensor network is asked to track a moving target which can only be viewed by a small number of sensor nodes [7] [8] [9]. This problem could also be encountered by robot flocking systems. The flocking system in which the robots cooperatively track a target is similar to a sensor network tracking a moving object. But in real robot flocking systems, the sensor range is limited and the robot camera view could be blocked by neighbours or obstacles. In such cases, only a small number of robots (leaders) can detect the moving targets and most robots (followers) have to predict or estimate the target position. Furthermore, the robots have to reach agreement about the target position in a distributed way.

It has been proved that the flocking system can keep cohesion by tracking a specific target (virtual leader) only if all the robots can get the target information [10]. But this method can not be applied to the leader-follower flocking system. In the leader-follower flocking system, only few robots (leaders) can “see” the target. The others (followers) do not know which individual is the leader and do not know where the target is. In this paper, we will investigate the target tracking of the leader-follower flocking system. In this system, the robots will use local vision sensors (pan/tilt cameras) to localise the target position. The robots who can detect the target will be considered as leaders. Otherwise, they are followers. The challenge of building such a system is how to reach an agreement about the target position among the robots. Firstly, the followers do not know where the target is; secondly, the leaders have different “opinions” about the target position because of observation errors. So a distributed consensus strategy is required to reach an agreement on the target position.

In this paper, the consensus algorithm is embedded into a Kalman filter and a consensus DKF (distributed Kalman filter) is developed for the cooperative target estimation. Each robot uses the Kalman filter to estimate the target position. The neighbour robots also interact with each other to share the target information and all members in the group gradually reach an agreement about the target position. The result of target estimation will be used for target tracking control. The fuzzy separation potential function introduced in the paper [12] will be used to achieve separation control. By combining the separation control and the target tracking control, a flocking controller will be developed. The performance of the flocking controller will be evaluated with realistic simulations.

## 2 Preliminary Knowledge

### 2.1 Robot Model

In this paper, we use the kinematic model of mobile robots. The kinematic model addresses the relationship between the position and the velocity of robots. In 2D space, a robot position can be expressed by a vector with 2 elements:

$$q_i = \begin{bmatrix} q_{ix} \\ q_{iy} \end{bmatrix} \quad (1)$$

where  $(q_{ix}, q_{iy})$  is the coordinate of a robot in 2D space.

In a flocking system with  $N$  robots, the state of these  $N$  robots can be represented by a position vector:

$$q = [q_1^T, q_2^T, \dots, q_i^T, \dots, q_N^T]^T \quad (2)$$

where  $q_i = [q_{ix}, q_{iy}]^T$  is the position (hand position) of robot  $i$ . The robot velocities can be described as

$$\dot{q} = [\dot{q}_1^T, \dot{q}_2^T, \dots, \dot{q}_N^T]^T \quad (3)$$

where  $\dot{q}$  is the derivative of  $q$  with respect to time. In our flocking controller,  $\dot{q}$  is considered as the output and  $q$  is considered as the input. Our purpose is to design a flocking controller with the following mathematical expression:

$$\dot{q} = f(q) \quad (4)$$

where  $f(q)$  is a function about the input “ $q$ ”.

In our flocking system, some of the robots (leaders) are asked to track a target. The state of the virtual leader or the target is also an input of the flocking controller. So the expression of leader-follower flocking controller should be written as:

$$\dot{q} = f(q, q_r) \quad (5)$$

where  $q_r$  is the state of the target.

### 2.2 Flocking Network Model

We assume that all robots in the flocking system have the same sensing or communication range. The communication range of robots is denoted as  $C$ . The robots consider the others with distance smaller than  $C$  as “neighbours”. To analyse the system, we build a mathematical model of a flocking network. The topology of a flocking network can be represented by a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  where  $\mathcal{V}$  is the set of vertices (robots) and  $\mathcal{E}$  is the set of edges (links between neighbours). The topology of a flocking network depends on distances between robots. A link only exists between a robot pair when their distance is smaller than  $C$ . If robot  $j$  is a neighbour of robot  $i$ , we have  $j \in N_i = \{j \in \mathcal{V}, j \neq i : \|q_i - q_j\| \leq C\}$  where  $N_i$  is the neighbour set of robot  $i$ .

### 2.3 Kalman Filter

In this paper, the flocking system is required to track the target which is represent by “ $q_r$ ”. Because a Kalman filter will be used to estimate the target position, it is necessary to introduce the related notations and basic knowledge.

As mentioned, the robots use local vision sensors to detect the target. Because of observation errors, the robots would get different results of the target position. The target observation result of robot  $i$  can be represented by  $z_{ir}$ . The observation results of all the robots can be written in a vector form:  $z_r = [z_{1r}^T, z_{2r}^T, \dots, z_{Nr}^T]^T$ . On the other hand, each robot uses a Kalman filter to estimate the target position. The estimated target position of robot  $i$  is denoted as  $\hat{x}_{ir} \in \mathbf{R}^4$ , which includes the estimates of target position and target velocity:

$$\hat{x}_{ir} = [\hat{q}_{ir}^T, \dot{\hat{q}}_{ir}^T]^T \tag{6}$$

where  $\hat{q}_{ir}$  is the estimate of the target position. The estimate vector of all robots is  $\hat{x}_r = [\hat{x}_{1r}^T, \hat{x}_{2r}^T, \dots, \hat{x}_{Nr}^T]^T \in \mathbf{R}^{4N}$ .

According to the definition of Kalman filter, the sensor model of robot  $i$  can be defined as:

$$z_{ir}(k) = Hx_r(k) + v(k); \quad p(v) \sim N(0, Q) \tag{7}$$

where  $Q$  is the covariance of observation error. We assume that the sensor models used by different robots are the same and the noise property is the same as well.

The estimation model of robot  $i$  can be written as:

$$\hat{x}_{ir}(k + 1) = A\hat{x}_{ir}(k) + K_i(k + 1)[z_{ir}(k) - H\hat{x}_{ir}(k)] \tag{8}$$

where  $K_i(k + 1)$  is the gain at step  $k$  and

$$K_i(k + 1) = P_i(k + 1)H^T[HP_i(k + 1)H^T + R]^{-1} \tag{9}$$

where  $P_i(k + 1)$  is the covariance of estimation error at step  $k$ .  $P_i$  is updated with equation:

$$P_i(k + 1) = AP_i(k)A^T + Q \tag{10}$$

### 3 Consensus Distributed Kalman Filter

In the flocking system, robots can reach an agreement on their aggregate information via consensus between adjacent robots. In this section, we will introduce how to embed the consensus algorithm into the Kalman filter and create a consensus DKF. The purpose is to reach an agreement about the estimation of the target position among the robots.

Firstly, we will introduce the discrete form of the alignment algorithm which is introduced in [11]:

$$\theta(k + 1) = (I - \mu\mathcal{L})\theta(k) \tag{11}$$

where  $\mathcal{L}$  is the Laplacian matrix of the flocking network.  $\theta$  is the vector of robot headings.  $\mu$  is a positive constant. Because the system (11) is a stable system, the system state  $\theta$  would asymptotically reach the equilibrium:  $\theta_1 = \theta_2 = \dots = \theta_N = \frac{1}{N} \sum_{i=1}^N \theta_i$ . As with the flocking centre estimation, a consensus algorithm of target estimation can be created by replacing the robot headings ( $\theta$ ) in (11) with the estimate of the target position of each robot:

$$\hat{x}_r(k + 1) = (I - \mu\hat{\mathcal{L}})\hat{x}_r(k) \tag{12}$$

where  $\hat{\mathcal{L}} = \mathcal{L} \otimes I_4$  is an 4D Laplacian matrix of the flocking network and  $\otimes$  is the Kronecker product. The outcome of the consensus algorithm should be  $\hat{x}_{1r} = \hat{x}_{2r} = \dots = \hat{x}_{Nr} = \frac{1}{N} \sum_{i=1}^N \hat{x}_{ir}$  which is the mean of the estimates.

According to the definition of the Laplacian matrix, the algorithm (12) can be written in a local form:

$$\hat{x}_{ir}(k + 1) = \hat{x}_{ir}(k) - \mu \sum_{j \in N_i} (\hat{x}_{ir}(k) - \hat{x}_{jr}(k)) \tag{13}$$

where  $N_i$  is the set of the neighbours of robot  $i$ . The robot  $i$  sends its estimate to all the neighbours via local communication, and all the neighbours send their estimates to robot  $i$  as well. Each robot uses the consensus algorithm to calculate its own estimate of the target position. Asymptotically, all the robots in the flocking system can reach an agreement on the estimate of the target position.

Assume that the estimate of the target position of robot  $i$  at step  $k$  is  $F_i(k)$  in the Kalman filter, which can be written as follows according to equation (8):

$$F_i(k) = A\hat{x}_{ir}(k) + K_i(k)[z_{ir}(k) - H\hat{x}_{ir}(k)] \tag{14}$$

By using the consensus algorithm, we have:

$$\hat{x}_{ir}(k + 1) = F_i(k) - \mu \sum_{j \in N_i} [F_i(k) - F_j(k)] \tag{15}$$

In the experiment, it is found that the value of  $\mu$  should be set around 0.3. If  $\mu$  is too small, the robots will take a long time to reach agreement which would result in considerable delay. Moreover,  $\mu$  should not be too large because it will influence the stability of the system and the robots will not reach agreement.

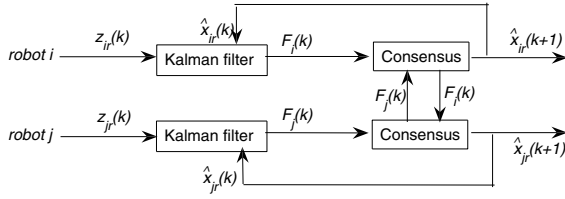


Fig. 1. Consensus DKF

Figure 1 shows the structure of consensus DKF between two neighbour robots. Each robot sends its observation results to the Kalman filter. The estimation results of the Kalman filter ( $F_i(k)$ ,  $F_j(k)$ ) will be sent to the consensus blocks. The two robots exchange the estimation result via the consensus blocks. The result of the consensus algorithm is the final target estimate of next step ( $\hat{x}_{ir}(k + 1)$ ,  $\hat{x}_{jr}(k + 1)$ ). Furthermore, these estimation results will be used to update the Kalman filter in the next step.

In general, the robots “negotiate” with their neighbours about the estimate of the target position at the current step. Although the followers do not have the target observation, they can estimate the target position via the consensus DKF. As a result, all the robots in the flocking system have estimates of the target position. In the next section, we will talk about how to use the target estimation to achieve tracking control.

### 4 Flocking Controller

The flocking controller consists of two components. One is the collision avoidance. Simply, adjacent robots should keep a specific distance. If the distance between two robots is too small, they attempt to separate. A fuzzy separation function has been developed in paper [12]. We will use the same fuzzy potential function to implement the separation control. The separation potential function between neighbours can be represented by  $H_s(\|q_i - q_j\|)$ . Another component is the tracking control that each robot uses to move towards the estimate of target position. We can use  $H_t(q_i, \hat{q}_{ir}) = \frac{1}{2}\|q_i - \hat{q}_{ir}\|^2$  to denote the tracking control potential function. So the flocking controller of robot  $i$  is designed as:

$$\dot{q}_i = -\nabla_{q_i} H_s(\|q_i - q_j\|) - k_r(q_i - \hat{q}_{ir}) \tag{16}$$

where  $k_r$  is a positive gain.

There are no differences between leader controllers and follower controllers in this flocking system. It means that leaders and followers are no longer predefined in the flocking system. Any robot who can detect the target is considered as a leader. The leaders would become followers if they lost their view of the target. So the robots are controlled by the same flocking controller and they autonomously change their roles (leader or follower) during the target tracking.

In the next step, we will use simulation to test the performance of the flocking system. The performance of the consensus DKF will also be evaluated via different parameters.

## 5 Simulations

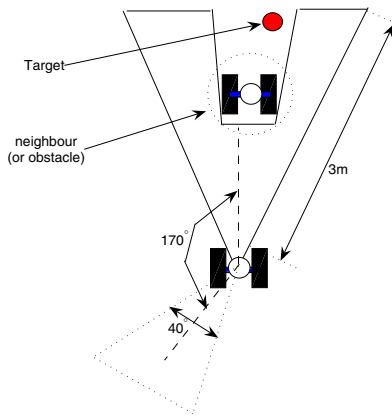
### 5.1 Camera Simulation

In simulation, the vision sensors are modeled as pan/tilt cameras. The sensing range of the camera is set as  $3m$ . The view of the camera is  $40^\circ$  and it can pan from  $-170^\circ$  to  $170^\circ$  to increase the view. The pan speed of the camera is  $0.8$  rad/s. The view of the camera could be blocked by neighbours or obstacles (as shown in figure 2). The camera will not stop scanning unless it detects a target. In figure 2, the target is located in the sensing range of the camera but one of the neighbours blocks the view. So the camera will continue scanning between  $-170^\circ$  and  $170^\circ$  until it finds the target.

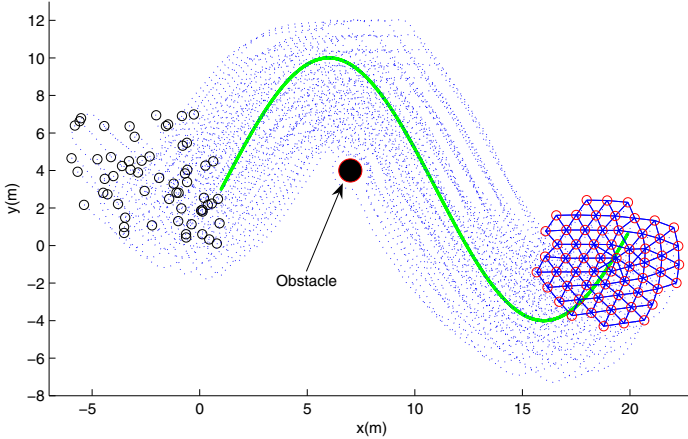
As with the real camera, the observation is imprecise and contains errors. In the simulator, the white Gaussian noise is used to simulate the errors of observation data. The covariance of the noise is set as the same covariance of the real camera observation .

### 5.2 Sine Trajectory Tracking

In the first simulation, the target moved along a sine shape trajectory. The sixty robots were randomly placed in a  $7m \times 7m$  area. The robot communication range was selected as  $C = 1.2m$  and the expected distance between robots was  $D = 1m$ . The robots used the consensus DKF to estimate the position of the target and used the flocking controller to flock and track the estimated target.



**Fig. 2.** View of camera in simulation



**Fig. 3.** Sixty robots tracking a sine trajectory

Figure 3 shows the stable tracking result. It illustrates that the flocking system was formed and smoothly tracked the target. The flocking network of the final state is plotted and shown in figure 3 as well. We can see that almost all the individuals in the flocking system maintained the specific distance with their neighbours. An obstacle was also placed near the target trajectory. It can be seen that the trajectories of the flocking system were slightly changed due to the obstacle avoidance manoeuvres. But the flocking system can still keep cohesion and pass the obstacle.

The process of sixty robots avoiding the obstacle is shown in figure 4. The first figure (figure 4(a)) shows the flocking system was moving to the obstacle. At  $t=115s$ , the flocking system just encountered the obstacle. Some robots detected the obstacle and moved around it. In figure 4(b), one of the robots was pushed away from the flocking system and it came back in figure 4(c). Finally, the flocking system passed the obstacle and continued the tracking. During the tracking, no robot collided with the obstacle and the flocking system kept cohesion.

As mentioned, the consensus DKF is used to estimate the target position. Figure 5(a) shows the results of the consensus estimation. The solid line denotes the true trajectory of the target; the dotted lines are the estimated results of the robots. At the beginning, robots did not reach an agreement about the target position. Especially at the first wave crest of the sine trajectory, the differences between the estimates were enormous. Gradually, all the estimates asymptotically approximated to a single trajectory which was close to the true trajectory. Figure 5(b) illustrates the mean error of target estimation. It can be seen that the mean error of target estimation fluctuated at the beginning of tracking. Gradually, it reached a relatively stable level of 0.5m.

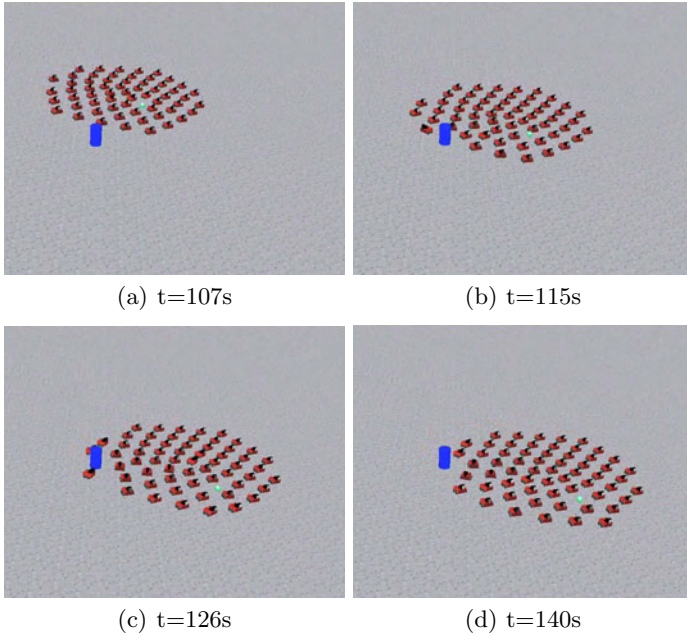


Fig. 4. Sixty robots tracking the target and avoiding the obstacle

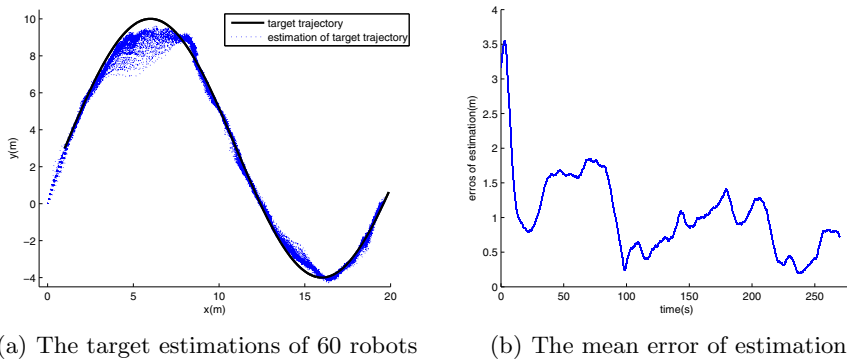


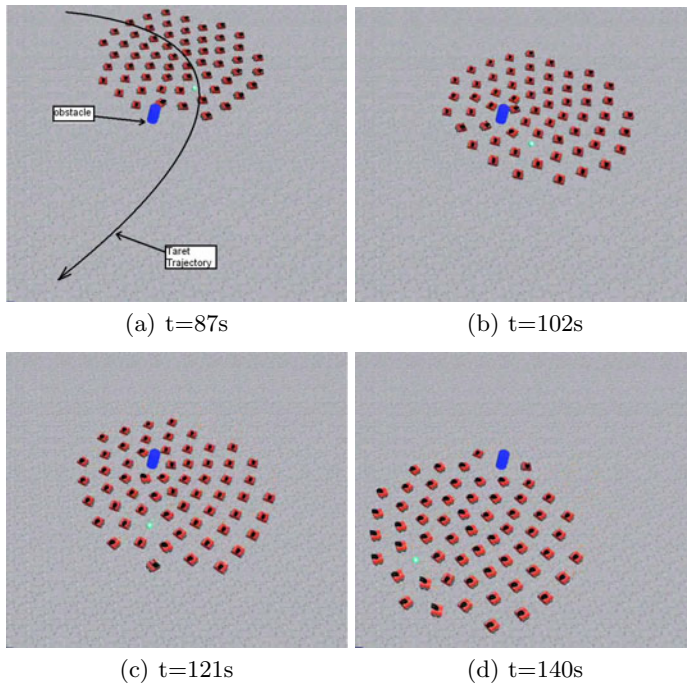
Fig. 5. Performance of target estimation

### 5.3 Circle Trajectory Tracking

In the second simulation, the target was moving along a circle trajectory with centre  $[0,-7]$  and radius  $8m$ . Figure 6 illustrates the trajectory tracking of the flocking system. It clearly shows that the flocking system was formed and smoothly tracked the target. It proves that the consensus DKF provided relatively accurate target information and the flocking controller can be used to track the target.



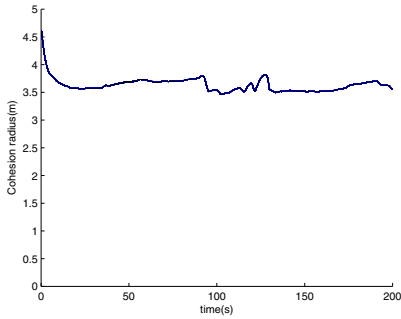
As with the last simulation, an obstacle was placed on the trajectory of the flocking system. At  $t=87s$ , the flocking system just encountered the obstacle. Some robots detected the obstacle and moved around it. Figure 6(b) and 6(c) show that the obstacle was in the middle of the flocking system. No robot collided with the obstacle and the flocking system was still cohesive. The flocking system successfully passed the obstacle and continued the target tracking at  $t=140s$  (6(d)).



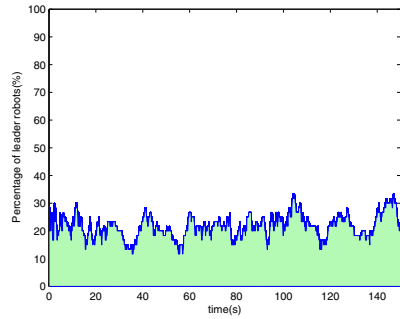
**Fig. 6.** Sixty robots track the target and avoid the obstacle

Figure 7(a) illustrates the cohesion radius of the flocking system during the tracking. It can be seen that the cohesiveness of the flocking system was slightly influenced during the obstacle avoidance (90s-140s). But the flocking system kept cohesion and the radius was stabilised around 3.7m.

There is no specific leader or follower in the flocking system. The robots who can detect the target are considered as leaders. The percentage of leaders in the flocking system is calculated and plotted in figure 7(b). The figure illustrates that about 20 percent of robots could detect the target during the whole process. It proves that the majority followers can be led by the minority leaders and that the flocking system can cohesively track the target.



(a) Cohesion radius of flocking system



(b) Percentage of leader robots in the flocking system

**Fig. 7.** performance evaluation

## 6 Summary

This paper presents a cooperative target tracking algorithm for leader-follower flocking systems. It uses a consensus DKF to establish the cooperation of target estimation. Each robot in the flocking system can benefit from the observation results of other robots and accordingly increases the target tracking performance. Meanwhile all robots in the group can use the agreed estimate of the target position to control their movements to flock cohesively. The agreement is achieved via a consensus algorithm embedded in the Kalman filter. Both the estimation of target position and flocking control are implemented in a distributed way.

A realistic simulator is used to test the large scale flocking system with sixty robots. The different parameters (cohesion radius, minimal distance, etc) are used to evaluate the performance of the flocking system. The simulation results show that the robots can cooperatively track the target and estimate a relatively accurate target position. With the help of fuzzy potential force, the flocking system is capable of avoiding obstacles and avoiding collision between neighbours. By investigating the percentage of leader robots inside the flocking system, we reach the conclusion that the flocking system can be led by few leaders who can see the target. The followers can also acquire the target position through the consensus DKF. All the robots benefit from the cooperative tracking and the flocking system keeps cohesion by tracking the estimated target position.

*Notes and Comments.* I would like to thank my supervisor, Doctor Dongbing Gu. I would not have been able to finish this paper without his suggestions and insights to help me solve challenging research problems. I would also like to thank Doctor Yanzhi Zhao and Doctor Tao Meng, for their invaluable advice and comments for this paper.

## References

1. Reynolds, C.W.: Flocks, herds, and schools: a distributed behavioural model. *Computer Graphics* 21(3), 25–34 (1987)
2. Egerstedt, M., Hu, X., Stotsky, A.: Control of mobile platforms using a virtual vehicle approach. *IEEE Trans. on Automatic Control* 46, 1777–1782 (2001)
3. Kalman, R.: A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* 82, 35–45 (1960)
4. Gu, D., Wang, Z.: Distributed regression over sensor networks: a support vector machine approach. In: *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3286–3291 (2008)
5. Bhansali, P., Gambini, S., Huang, P., Roychowdhury, J.: Networks of oscillators for synchronization. In: *Proc. of the GSRC Annual Symposium* (2008)
6. Carli, R., Chiuso, A., Schenato, L., Zampieri, S.: Distributed Kalman filtering using consensus strategies. In: *Proc. of the 44th IEEE Conference on Decision and Control* (2007)
7. Gupta, V., Chang, T., Hassibi, B., Murray, R.: On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. *Automatica* 44, 1295–1303 (2008)
8. Medeiros, H., Park, J., Kak, A.: Distributed object tracking using a cluster-based Kalman filter in wireless camera networks. *IEEE Journal of Selected Topics in Signal Processing* 2, 448–463 (2008)
9. Khan, U., Moura, J.: Distributing the Kalman filter for large-scale systems. *IEEE Trans. on Signal Processing*, 4919–4935 (2008)
10. Olfati-Saber, R.: Flocking for multi-agent dynamic systems: algorithms and theory. *IEEE Trans. on Automatic Control* 51(3), 401–420 (2006)
11. Jadbabaie, A., Lin, J., Morse, S.A.: Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Trans. on Automatic Control* 48(6), 988–1001 (2003)
12. Gu, D., Wang, Z.: LeaderDFollower Flocking: Algorithms and Experiments. *IEEE Transactions on Control Systems Technology* 17, 1211–1219 (2009)

# A Decentralized Junction Tree Approach to Mobile Robots Cooperative Localization

Hua Mu, Meiping Wu, Hongxu Ma, and Wenqi Wu

College of Mechatronic Engineering and Automation,  
National University of Defense Technology, Hunan 410073, P.R. China  
huamu08@gmail.com

**Abstract.** This paper presents a decentralized solution to the cooperative localization of mobile robot teams. The problem is cast as inference on a dynamic Bayesian network (DBN) of Gaussian distribution, which is implemented incrementally by decomposing the DBN into a sequence of chain graphs connected by the interfaces. The proposed inference scheme can make use of the sparsity of the chain graphs and achieve efficient communication. In our decentralized formulation, the local sensor data at each robot are organized as potentials of the cliques of junction trees; message passing between robots updates the clique potentials to realize information sharing. Each robot can get optimal estimates of its own states. The method is optimal in the sense that it makes no approximations apart from the usual model liberalization. The performance of the proposed algorithm is evaluated with simulation experiments.

**Keywords:** Cooperative localization, decentralized inference, dynamic Bayesian network, chain graph, interface, junction tree.

## 1 Introduction

Multiple mobile robot systems have a wide range of potential applications, such as search and rescue [1], exploration [2,3] and surveillance [4]. *Cooperative localization* is an efficient means to estimate the joint state of all the robots by sharing information among the team. The robots may be correlated by inter-robot observations or the same external landmarks employed for self-localization. This paper investigates the former case, which is more general since a robot itself can be regarded as a mobile “landmark” for other robots.

A centralized solution to the cooperative localization problem is straightforward yet requires a processing center with high capability. It is also fragile since the entire system fails if the processing center fails. A decentralized solution, however, does not rely on any single processing center and the computation is distributed among the robots.

Early decentralized solutions to cooperative localization, such as maximum likelihood over a finite time-horizon [5] or Monte Carlo filters [6], made the assumption that the trajectory estimate from one platform is uncorrelated to the estimate of another; an assumption that eventually causes over-confidence

and inconsistency in the estimation process. Consistent distributed estimation built on the EKF has been demonstrated for cooperative localization with indoor mobile robots [7]. Cross-correlation terms of the joint covariance matrix are distributed among the platforms after each update step, which is an operation that scales poorly with the number of robots since each update of an observation revises every term of the joint covariance matrix. The update of the observations made at the same time must be implemented sequentially, which will cause long communication delay. A different EKF-based approach had each platform maintain a bank of EKFs [8]. This scheme is computationally expensive as the total number of filters grows exponentially with the size of the team, and suboptimal as the book-keeping strategy permits fusion of only a subset of the available information.

A recent proposal [9] employed the canonical form (or information form) parameters to solve cooperative localization, where each platform first estimates its motion using only local sensor data, then shares its information across the network using an algorithm that employs a distributed Cholesky modification. There is no processing center, yet each platform accumulates the historical information carried with Cholesky factors and the communication is expensive.

This paper proposes a decentralized solution, where the local sensor data at each robot are organized as potentials of cliques of a junction tree and the information is shared by message passing among the team. The removal of the processing center allows load balance among the robots and team reorganization in the face of robot failure. The incremental inference scheme presented in the paper makes use of the sparsity of the problem and reduces the communication costs compared with the conventional interface algorithms. Our decentralized solution is based on the *exact* inference algorithm – junction tree algorithm [10][11]; it makes no approximations apart from the usual model liberalization; once all of the available information has propagated through the system, it will generate the same estimates as a centralized estimator. Thus the proposed solution is an *optimal* estimator. The solution is verified and analyzed with simulation experiments.

One influence to this work is the concept of interface developed in the forward interface algorithm [12] for online inference on DBNs. The interface supports the inference on DBNs to be decomposed onto a sequence of sub-networks; each sub-network is static and can be solved using junction tree algorithms. The interface is forced to form a clique when constructing junction trees for a sub-network in the forward interface algorithm. We remove this constraint by introducing the concept of chain graph and the operation of PUSH to the incremental inference algorithm design.

Another influence to this work is the application of the junction tree algorithms to some distributed dynamic problems. The problems of networked cameras calibration by tracking a moving target [13] and temperature monitoring by a sensor network [14] are solved in an approximate distributed probabilistic inference architecture, where two layers of junction tree are required: an external junction tree for message passing and a network junction tree for allocation of

the variables among the sensor network. In [15] the multi-robot localization and mapping problem is solved using a distributed multifrontal QR factorization approach. With static landmarks and without the inter-robots measurements, the constructed junction tree is claimed to have a nice structure, wherein the cliques composed of the landmarks seen by one robot alone form a subtree and are descendants of the cliques containing landmarks seen by multiple robots. The cooperative localization problem, where the sensors and the targets are all moving and correlated to each other, is more complex than the above applications.

## 2 Problem Formulation

Consider a team of robots moving in a 2D environment. Each robot is equipped with dead-reckoning sensors to measure self-motion; odometer or inertial measurement unit for example. Some robots carry sensors that provide relative position measurements (range and/or bearing) among robots or between robots and the environment; such as cameras, laser range-finders. Some robots can correct its dead-reckoning error. In the case that the environment map is known, the measurements between a robot and the environment can be utilized to bound the dead-reckoning error. Otherwise, the sensor like GPS can provide absolute localization information. The robots can communicate with each other. The cooperative localization task is to estimate the state (position, orientation, velocity etc) of each robot making use of both its own observations and those observations made by and of other robots. The robots are identified by capitals  $A, B, C, \dots$ .

### 2.1 State Space Model

The state of robot  $A$  at time step  $k$  is denoted  $\mathbf{x}_k^A$ . Assume the motion model of robot  $A$  is given as

$$\mathbf{x}_{k+1}^A = f(\mathbf{x}_k^A, \mathbf{u}_k) + G_k \mathbf{w}_k, \quad (1)$$

where  $\mathbf{u}_k$  is the system input at time step  $k$ ,  $\mathbf{w}_k$  is the process noise,  $\mathbf{w}_k \sim \mathcal{N}_m(\mathbf{w}_k; \mathbf{0}, Q)$ , and  $G_k$  is a matrix with proper dimensions.  $\mathcal{N}_m(\mathbf{v}; \mu, P)$  represents a Gaussian distribution over  $\mathbf{v}$  with mean  $\mu$  and covariance  $P$ .

The GPS-like measurement of robot  $A$  at time step  $k$  is denoted  $\mathbf{z}_k^A$ . The observation model is given as

$$\mathbf{z}_k^A = h_1(\mathbf{x}_k^A) + \mathbf{r}_{1k}, \quad (2)$$

where  $\mathbf{r}_{1k}$  is the measurement noise,  $\mathbf{r}_{1k} \sim \mathcal{N}_m(\mathbf{r}_{1k}; \mathbf{0}, R_1)$ .

The measurement that robot  $A$  makes to robot  $B$  at time step  $k$  is denoted  $\mathbf{z}_k^{AB}$ . The observation model is defined as

$$\mathbf{z}_k^{AB} = h_2(\mathbf{x}_k^A, \mathbf{x}_k^B) + \mathbf{r}_{2k}, \quad (3)$$

where  $\mathbf{r}_{2k}$  is the measurement noise,  $\mathbf{r}_{2k} \sim \mathcal{N}_m(\mathbf{r}_{2k}; \mathbf{0}, R_2)$ .

Since the functions  $f(\cdot)$ ,  $h_1(\cdot)$  and  $h_2(\cdot)$  are generally nonlinear, Taylor expansions are used to linearize the models to make the inference tractable. For

a nonlinear  $f(\cdot)$  in (11), the Taylor expansion is developed around  $\mu_k^A$ , the best available mean estimate of  $\mathbf{x}_k^A$ ,

$$\mathbf{x}_{k+1}^A = f(\mu_k^A, \mathbf{u}_k) + F(\mathbf{x}_k^A - \mu_k^A) + G_k \mathbf{w}_k, \tag{4}$$

where  $F = \left. \frac{\partial f}{\partial \mathbf{x}_k^A} \right|_{\mathbf{x}_k^A = \mu_k^A}$ .

The nonlinear observation model (2) can be linearized as

$$\mathbf{z}_k^A = h_1(\mu_k^A) + H(\mathbf{x}_k^A - \mu_k^A) + \mathbf{r}_{1k}, \tag{5}$$

where  $H = \left. \frac{\partial h_1}{\partial \mathbf{x}_k^A} \right|_{\mathbf{x}_k^A = \mu_k^A}$ .

Similarly, the nonlinear observation model (3) is linearized as

$$\mathbf{z}_k^{AB} = h_2(\mu_k^A, \mu_k^B) + H_A(\mathbf{x}_k^A - \mu_k^A) + H_B(\mathbf{x}_k^B - \mu_k^B) + \mathbf{r}_{2k}, \tag{6}$$

where  $H_A = \left. \frac{\partial h_2}{\partial \mathbf{x}_k^A} \right|_{\mathbf{x}_k^A = \mu_k^A}$  and  $H_B = \left. \frac{\partial h_2}{\partial \mathbf{x}_k^B} \right|_{\mathbf{x}_k^B = \mu_k^B}$ .

### 2.2 Gaussian DBNs

A DBN is a directed acyclic graph with an infinite set of nodes; each node of the graph corresponds one-to-one with a vector of variables  $\mathbf{v} \in \mathcal{V}$  and associates with a conditional probability distribution  $P(\mathbf{v}|pa(\mathbf{v}))$ , where  $pa(\mathbf{v})$  is the set of variables corresponding to the parents of the node representing  $\mathbf{v}$  in the graph. Without losing clarity, the symbol  $\mathbf{v}$  refers both a variable and its corresponding node of a DBN throughout this paper. A DBN where each conditional distribution  $P(\mathbf{v}|pa(\mathbf{v}))$  is a Gaussian distribution is a Gaussian DBN.

Assume the initial state of each robot is a Gaussian variable. The linearized system (4-6) defines the following conditional distributions

$$\begin{aligned} P(\mathbf{x}_{k+1}^A | \mathbf{x}_k^A) &= \mathcal{N}_m(\mathbf{x}_{k+1}^A; f(\mu_k^A, \mathbf{u}_k) - F\mu_k^A + F\mathbf{x}_k^A, \bar{Q}), \\ P(\mathbf{z}_k^A | \mathbf{x}_k^A) &= \mathcal{N}_m(\mathbf{z}_k^A; h(\mu_k^A) - H\mu_k^A + H\mathbf{x}_k^A, R_1), \\ P(\mathbf{z}_k^{AB} | \mathbf{x}_k^A, \mathbf{x}_k^B) &= \mathcal{N}_m(\mathbf{z}_k^{AB}; h_2(\mu_k^A, \mu_k^B) - H_A\mu_k^A - H_B\mu_k^B + H_A\mathbf{x}_k^A + H_B\mathbf{x}_k^B, R_2), \end{aligned} \tag{7}$$

where  $\bar{Q} = G_k Q G_k^T$ . Then the joint distribution of the state variables and the observation variables can be represented by a Gaussian DBN, whereby each node represents a state variable or an observation variable and the conditional probability distributions are provided by (7).

### 3 Incremental Inference on Gaussian DBNs

The probabilistic inference task is to compute the marginals over some nodes given some other nodes taking fixed values. The incremental inference schemes

that have been developed most are concerned with DBNs with discrete variables. We focus on Gaussian DBNs and adopt the canonical form parameterizations for Gaussian distributions following [10]. It will be shown that the canonical form parameterizations makes the computation involved in the inference simple.

### 3.1 Subgraph Construction

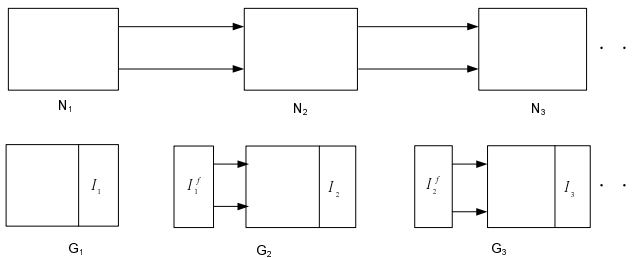
The incremental inference for a dynamic system can be realized by implementing inference for a sequence of *static* subsystems; the two consecutive subsystems overlap partially for information inheritance. The overlapped part is the *interface*, which is formally defined as follows.

*Definition.* Divide a DBN into a sequence of BNs  $N_1, N_2, \dots$ , where two consecutive ones are connected by temporal edges. The interface  $\mathcal{I}_k$  is defined to be the set of nodes in  $N_k$  which have children in  $N_{k+1}$  ( $k = 1, 2, \dots$ ).

Let  $V_{k+1}$  denote the union of the nodes of  $\mathcal{I}_k$  and the nodes in  $N_{k+1}$ . The  $(k+1)^{th}$  subsystem defines  $P(V_{k+1})$ , the marginal distribution over  $V_{k+1}$ . The graphical model  $G_{k+1}$  for  $P(V_{k+1})$  can be constructed incrementally as follows.

- Eliminate  $V_k \setminus \mathcal{I}_k$  from  $G_k$  and the resultant graph is  $I_k^f$ .
- Connect  $N_{k+1}$  to the graph  $I_k^f$ ;

Figure 1 illustrates the construction of the graphical models  $G_1, G_2, \dots$ .



**Fig. 1.** (Top) A DBN is separated into a sequence of linked BNs  $N_k$ . (Bottom) The  $(k+1)^{th}$  subsystem is defined over the variables in  $V_{k+1} \triangleq \mathcal{I}_k \cup N_{k+1}$  and the corresponding graphical model is denoted  $G_{k+1}$ . The subgraph induced by the nodes  $\mathcal{I}_k$  is denoted  $I_k$ . The graph after eliminating  $V_k \setminus \mathcal{I}_{k-1}$  from  $G_k$  is denoted  $I_k^f$ .  $G_{k+1}$  is obtained by linking  $I_k^f$  and  $N_{k+1}$ .  $G_1 = N_1$ .

Since the graph  $I_k^f$  quickly becomes a clique with graph elimination process,  $G_{k+1}$  can be built without the actual graph elimination operations. Instead,  $G_{k+1}$  is built by forming a clique with the nodes  $\mathcal{I}_k$  and augmenting  $N_{k+1}$ .

The inference is implemented on each  $G_k$  incrementally. The graphical model  $G_k$  for sub-networks of a DBN constructed as above is a chain graph. The sub-routine of inference on chain graphs is presented in the following subsection.



### 3.2 Inference on Chain Graphs

*Chain graph* is a kind of graphical model that has no directed cycles [11]. Given a chain graph  $G$  with the node set  $V$ , after removing all the directed edges of  $G$ , each connected component of the resultant graph is a *chain component* of  $G$ , denoted  $M_1, M_2 \cdots M_n$ . It holds that  $\bigcup_{i=1}^n M_i = V$ . Any distribution  $P$  over  $V$  satisfying the Markov property of  $G$  has the following factorization form

$$P(V) = \prod_{i=1}^n P(M_i | pa(M_i)), \tag{8}$$

where  $pa(M_i)$  is the set of parents of  $M_i$  in  $G$ .

**Junction Tree Initialization.** The junction tree for a given chain graph is constructed in the usual way: moralization, triangulation and minimum spanning tree formation. See for example [16, 11] for details. Here assume we are at the point where we have constructed a junction tree  $\mathcal{T}$  with cliques  $\mathcal{C} = \{C_1, C_2, \dots\}$ . The separator associated with two adjacent cliques,  $C_i$  and  $C_j$ , is denoted as  $S_{ij} = C_i \cap C_j$ . To initialize a junction tree, the probability factors in (8) are assigned to the cliques. A clique potential  $\psi_C$  is the product of the factors assigned to it; the potential is unity if there is no factors assigned. All the separator potentials are initialized as unity.

For the chain graph  $G_{k+1}$  generated from Gaussian DBNs as described in Section 3.1, there are two types of chain components. One is the chain component  $I_k^f$ , the other is each node in  $N_{k+1}$ . The former type of chain component has no parents and  $P(I_k^f)$  follows Gaussian distribution. The latter type of chain component is linearly related to its parents, which can be seen from (4-6),

$$M_i = T \cdot pa(M_i) + \mathbf{u}, \tag{9}$$

where  $\mathbf{u} \sim \mathcal{N}_m(\mathbf{u}; \bar{u}, U)$ .  $P(M_i | pa(M_i))$  is a Gaussian distribution with the canonical form parametrization as

$$\mathcal{N}_I \left( \begin{bmatrix} pa(M_i) \\ M_i \end{bmatrix}; \begin{bmatrix} -T^T U^{-1} \bar{u} \\ U^{-1} \bar{u} \end{bmatrix}, \begin{bmatrix} T^T U^{-1} T & -T^T U^{-1} \\ -U^{-1} T & U^{-1} \end{bmatrix} \right), \tag{10}$$

where  $\mathcal{N}_I(\mathbf{v}; \xi, \Omega)$  represents a Gaussian distribution over  $\mathbf{v}$  with information vector  $\xi$  and information matrix  $\Omega$ .

Consider the chain graph in Fig. 2(a) which is a graphical model of two robots  $A$  and  $B$  in a time interval. A junction tree constructed for the chain graph is shown in Fig. 2(b).

**Evidence Incorporation.** When a node  $\mathbf{v}$  takes a fixed value  $\mathbf{v} = v$ , any clique whose domain contains  $\mathbf{v}$  is required to update its potential to be consistent with the evidence. Evidence incorporation is a conditioning operation. Partition the nodes in a clique into two parts,  $C = \{\mathbf{v}_1, \mathbf{v}_2\}$ , with the evidence  $\mathbf{v}_2 = v_2$ . Given the clique potential

$$\psi_C = \mathcal{N}_I \left( \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix}; \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}^T & \Omega_{22} \end{bmatrix} \right), \tag{11}$$

after incorporating  $\mathbf{v}_2 = v_2$ , we have

$$\psi_C^* = \mathcal{N}_I(\mathbf{v}_1; \xi_1 - \Omega_{12}v_2, \Omega_{11}). \tag{12}$$

See for example [12] (Page 149) for the derivation.

**Message Passing.** The standard Hugin message passing strategy is employed. Let  $C_i$  and  $C_j$  be two adjacent cliques. Passing a message from  $C_i$  to  $C_j$  updates the potentials as

$$\psi_{S_{ij}}^* = \int \psi_{C_i} d(C_i \setminus S_{ij}) \quad \psi_{C_j}^* = \psi_{C_j} \frac{\psi_{S_{ij}}^*}{\psi_{S_{ij}}} \tag{13}$$

where the asterisk means “updated value of”.

Each message passing involves a marginalization and rescaling of the potential of the receiver. The potentials multiplication and division in the rescaling are realized by the summation and subtraction of the potential parameters; the computation simplicity is apparent.

When the messages are passed along every edge in both directions in proper schedule, the clique and separator potentials are marginals of  $P(\mathcal{V})$  over the domains.

**Junction Tree Rearrangement.** To calculate the marginal over a set of nodes which are not contained in a clique, we can rearrange the junction tree employing the PUSH operations to create a clique containing all the nodes in question. Consider two adjacent cliques  $C_i$  and  $C_j$ . A group of variables  $V \subseteq (C_i \setminus S_{ij})$  can be PUSHed to  $C_j$  as following:

- The clique  $C_j$  is extended to  $C_j^* = C_j \cup V$ , and similarly  $S_{ij}^* = S_{ij} \cup V$ ;
- The potentials are updated as

$$\psi_{S_{ij}^*} = \int \psi_{C_i} d(C_i \setminus S_{ij}^*), \quad \psi_{C_j^*} = \psi_{C_j} \frac{\psi_{S_{ij}^*}}{\psi_{S_{ij}}}$$

When calculating  $P(V)$  where the nodes in  $V$  are in different cliques, first form the smallest connected subtree that contains all the nodes in  $V$ , and then use PUSH operations to make  $V$  contained in a single clique  $C$ . The marginal  $P(V)$  can be computed from  $\psi_C^*$ .

### 3.3 Algorithm Overview

The entire algorithm for incremental inference on a DBN is as follows.

1. For  $k = 1$ ,  
 $G_1 = N_1$ . Identify the interface  $\mathcal{I}_1$ . Construct  $\mathcal{T}_1$  for  $G_1$ . Implement inference on  $\mathcal{T}_1$  and calculate  $P(\mathcal{I}_1)$ . Eliminate  $V_1 \setminus \mathcal{I}_1$  from  $G_1$  and the resultant graph is denoted  $I_1^f$ .

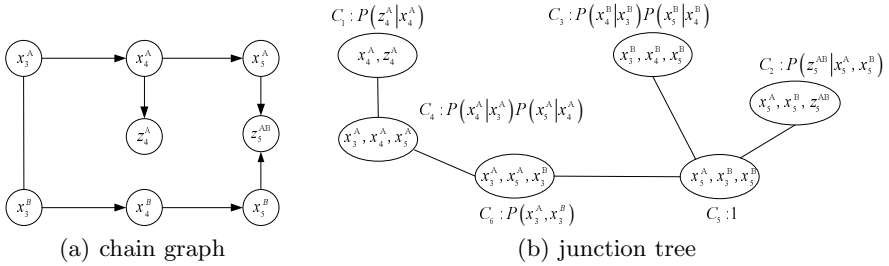


Fig. 2. A chain graph and one of its initialized junction tree

2. For  $k \geq 2$ ,
  - (a) Construct  $G_k$  by connecting  $N_k$  to the graph  $I_{k-1}^f$ ;
  - (b) Construct  $\mathcal{T}_k$  for  $G_k$ ; the probability factors associated with  $\mathcal{T}_k$  are from two parts, the first is  $P(\mathcal{I}_{k-1})$  which is provided by the inference on  $\mathcal{T}_{k-1}$ , the second is the conditional probability of each node in  $N_k$ ;
  - (c) Implement inference on  $\mathcal{T}_k$ . Identify  $\mathcal{I}_k$  and calculate  $P(\mathcal{I}_k)$  to pass to  $\mathcal{T}_{k+1}$ .
  - (d) Eliminate  $V_k \setminus \mathcal{I}_k$  from  $G_k$  and the resultant graph is  $I_k^f$ .

Recall the forward interface algorithm in [12] which forces both  $\mathcal{I}_k$  and  $\mathcal{I}_{k+1}$  to form a clique when constructing  $\mathcal{T}_{k+1}$ .  $\mathcal{I}_k$  formed in a clique actually is the result of the node elimination of  $V_k \setminus \mathcal{I}_k$ . However, it is not necessary to form  $\mathcal{I}_{k+1}$  into a clique for the calculation of  $P(\mathcal{I}_{k+1})$ ; we can rearrange the junction tree to create a clique containing  $\mathcal{I}_{k+1}$ . Thus the forward interface algorithm is a special implementation of the general algorithm proposed in the paper. The constraint of forming  $\mathcal{I}_{k+1}$  in a clique of  $\mathcal{T}_{k+1}$  may sharply reduce the sparsity of the graph  $G_{k+1}$  and result in unnecessary large cliques. By removing this constraint, the inference can be implemented on a junction tree with cliques as small as possible and the inference costs could be reduced.

## 4 A Decentralized Solution to Cooperative Localization

We apply the incremental inference algorithm to the problem of mobile robots cooperative localization in a decentralized fashion, eliminating the need of a highly capable processing center.

### 4.1 Overview of the Decentralized Solution

In our decentralized implementation, the cliques of junction trees are assigned to robots and the communication among the robots is guided by *link tree*. Each clique of a junction tree is assigned to a robot and labeled with the robot ID; by combining the neighboring cliques with the same robot ID into a node, the resultant tree is link tree. The link tree bridges the gap between the clique oriented inference algorithm and the robot oriented algorithm. Figure 3 gives a

link tree for the junction tree in Fig. 2(b) with cliques  $C_1, C_4, C_6$  assigned to Robot  $A$  and  $C_2, C_3, C_5$  to Robot  $B$ .

In our implementation, the *symbolic* global information is processed at a coordinator robot and the numerical information is calculated and communicated between the robots. In the inference on a subsystem  $G_k$ , the coordinator robot processes the following symbolic information.

- Build the chain graph  $G_k$ .
- Construct junction tree  $\mathcal{T}_k$  for  $G_k$ . Note it is the symbolic potentials that are assigned to cliques.
- Assign the cliques to robots and form a link tree.
- Identify the subtree to calculate  $P(\mathcal{I}_k)$  if the nodes  $\mathcal{I}_k$  are not contained in a single clique.

The overall inference on  $G_k$  proceeds as following:

1. Each robot calculates local probability factors  $P(\mathbf{x}_{k+1}^A | \mathbf{x}_k^A)$ ,  $P(\mathbf{z}_k^A | \mathbf{x}_k^A)$  and  $P(\mathbf{z}_k^{AB} | \mathbf{x}_k^A, \mathbf{x}_k^B)$  from local sensor measurements according to (10).
2. The coordinator robot processes the global information for the time interval  $[t_{k-1}, t_k]$  at time step  $t_h$ . Note  $t_k < t_h$  to facilitate the delay caused by communications among the robots. The coordinator transmits the processed information to relevant robots.
3. On receiving  $\mathcal{T}_k$  and the assignment information, each robot calculates the numerical potential for each clique assigned to it.
4. The robots implement message passing cooperatively.

## 4.2 Implementation Issues

To construct  $G_k$ , the coordinator collects the symbolic information of state evolvments and observations from all the robots in the team, e.g., the symbolic information for a range measurement made by robot  $B$  to robot  $A$  at time step  $k$  is  $\{\mathbf{r}, (B, A), k\}$ .  $G_k$  is obtained by connecting  $I_{k-1}^f$  with the BN of the state and observation variables in the time interval  $[t_{k-1}, t_k]$ .

To achieve a junction tree topology that makes decentralized implementation efficient, we adopt the following elimination ordering for triangulation: first the observation variables, then the state variables without inter-robot measurements and finally the states with inter-robot measurements. The fill-ins during triangulation can be further reduced by optimizing the elimination ordering inside the group of states with inter-robot measurements. This leads to a junction tree where an observation variable appears only once in a leaf clique and the cliques composed of the states without inter-robot measurements of one robot form a connected subtree. The junction tree built in this way can make full use of the locality of computation and reduce the computation and communication costs.

The potential initialization and the assignment of cliques to robots are both aimed at reducing communication requirements between robots. A clique without symbolic factors could be assigned to any robot; we assign it to one of its neighbors in the link tree to reduce the latent communication in message passing.

## 5 Experiments

Simulation experiments are devised for two purposes, solution verification and communication costs analysis. In realistic decentralized systems, communication usually consumes significant energy and is one of a key factors for system design.

### 5.1 Verification

The verification of the proposed algorithm is carried out in a simulation scenario of four mobile robots (labeled with  $A$ ,  $B$ ,  $C$  and  $D$ ) which can communicate with each other. Assume each robot has odometer for motion measurement. To demonstrate the information sharing among robots, the other sensing is limited as follows. Robot  $A$  has a GPS receiver. It also has a laser to make range-bearing measurements to other robots. Robot  $C$  measures its yaw with a gyro. It does not make any exteroceptive measurements. Robot  $B$  and  $D$  use their own laser to make range-bearing measurements to other robots. Data association for the range-bearing observations is assumed known.

The two-wheel vehicle model is used for the simulation. Among all the simulation parameters, the initial error is chosen be small enough to mitigate the problem of large non-linearities, and the other parameter values are non-critical.

Figure 4 provides the robot tracks with and without cooperative localization, together with the ground truth. The tracks given by cooperative localization and the ground truth match together rapidly in the case of small initial errors. The cooperative localization improves the localization accuracy for every robot, especially for Robot  $B$  and Robot  $D$ .

### 5.2 Costs Analysis

This section analyzes the computation and communication costs of the solution proposed in Section 4, compared with the solution utilizing the forward interface algorithm. The main costs in the decentralized solutions lie in the message



Fig. 3. A link tree for the junction tree in Fig. 2(b)

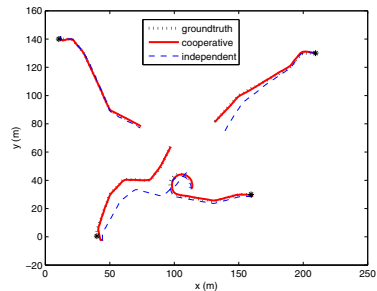


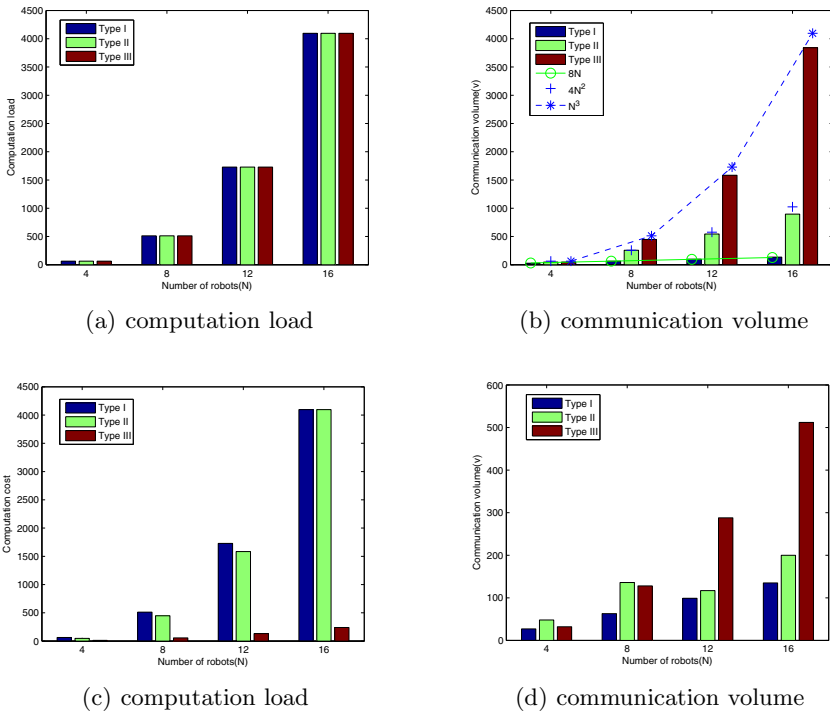
Fig. 4. The robot tracks. The starting points are labeled with asterisks.

passing in the junction tree algorithm. Consider the representative subnetworks composed of two consecutive time slices where the state nodes form a chain component and the inter-robot measurements in the recent time slice take on the following types of “sparsity”.

- Type I.** The inter-robot measurements are made between a leader robot and all the other robots.
- Type II.** Order the robots in a ring. The inter-robot measurements are made only between a robot and its neighbors.
- Type III.** The inter-robot measurements are made between any two robot.

The numbers of robots investigated are  $N = 4, 8, 12, 16$ . The inference task is to calculate the marginal over the interface nodes, i.e., the state nodes in the recent time slice. Note the localization accuracy resulted from different measurement types is not our concern.

Consider a message which contains the information matrix and the information vector of a Gaussian probability defined over  $m$  states. The volume of communicating such a message is counted as  $m^2$ . This is a relative measure since it is irrelevant with the state dimension. The computation load is measured with the number of multiplication operations during marginalization. Figure 5 shows



**Fig. 5.** The total costs (a and b) and the maximum single robot costs (c and d) of doing inferences for representative subnetworks by proposed solutions

the costs to do the inferences by the proposed solution, wherein (a) and (b) are the total computation load and communication volume respectively, and (c) and (d) are the maximum single robot costs.

The costs of the decentralized solution utilizing the forward interface algorithm can be inferred from the above results. The total computation load is determined by the marginalization operation, which is independent of the construction of the junction trees for a given representative subnetwork. Thus total computation loads of the forward interface algorithm can also be depicted by Fig. 5(a). The total communication volume, however, is related to the junction tree construction. Since the junction tree of the representative subnetworks with different measurement types constructed in the forward interface algorithm is the same with that in our solution for Type III measurements with the same team size, the total communication volumes are the same with our solution for Type III measurements for a given team size. This demonstrates the communication reduction by exploring the sparsity of the graphical models with the proposed solution. The maximum single robot cost of the decentralized solution utilizing forward interface algorithm is independent of the measurement types and the same with our solution for Type III measurements for a given team size.

## 6 Conclusions

This paper presents a decentralized algorithm for cooperative localization which is efficient and optimal. In our formulation, the local sensor information at each robot is processed as clique potentials and then shared among the team through message passing. The fully decentralized structure requiring no processing center provides load balance among the robots and the reorganization ability in the face of robot failure. The communication cost analysis reveals how is the cost related to the team size and the inter-robot measurements, which is meaningful to the design of realistic systems.

In the future, we intend to implement the proposed algorithm on some Pioneer indoor robots for further verification and analysis. We also expect to further reduce the communication cost in the general scenarios where any robot in the team can observe any other at all times. By introducing some approximate techniques to our incremental inference algorithm, the decentralized solution is expected to enjoy the cost reduction while keeping an acceptable accuracy.

## References

1. Sugiyama, H., Tsujioka, T., Murata, M.: Collaborative movement of rescue robots for reliable and effective networking in disaster area. In: 2005 International Conference on Collaborative Computing: Networking, Applications and Worksharing (2005)
2. Grabowski, R., Navarro-Serment, L.E., Paredis, C.J.J., Khosla, P.K.: Heterogeneous teams of modular robots for mapping and exploration. *Autonomous Robots* 8(3), 293–308 (2000)

3. Burgard, W., Moors, M., Stachniss, C., Schneider, F.: Coordinated multi-robot exploration. *IEEE Transactions on Robotics and Automation* 21(3), 376–386 (2005)
4. Merino, L., Caballero, F., de Dios, J.M., Ferruz, J., Ollero, A.: A cooperative perception system for multiple UAVs: Application to automatic detection of forest fires. *Journal of Field Robotics* 23(3-4), 165–184 (2006)
5. Howard, A., Matari, M., Sukhatme, C.: Localization for mobile robot teams: A distributed MLE approach. In: Siciliano, B., Dario, P. (eds.) *Experimental Robotics VIII*, pp. 146–155. Springer, Heidelberg (2003)
6. Fox, D., Burgard, W., Kruppa, H., Thrun, S.: A probabilistic approach to collaborative multi-robot localization. *Autonomous Robots* 8(3), 325–344 (2000)
7. Roumeliotis, S., Bekey, G.: Distributed multi-robot localization. *IEEE Transactions on Robotics and Automation* 18(5), 781–795 (2002)
8. Bahr, A., Walter, M., Leonard, J.: Consistent cooperative localisation. In: *IEEE International Conference on Robotics and Automation* (2009)
9. Mu, H., Baily, T., Thompson, P., Durrant-Whyte, H.: Decentralised solutions to the cooperative multi-platform navigation problem. *IEEE Transactions on Aerospace and Electronic Systems* (2010) (in press)
10. Lauritzen, S.: Propagation of probabilities, means and variances in mixed graphical association models. *Journal of the American Statistical Association* 87, 1098–1108 (1992)
11. Cowell, R.G., Dawid, A.P., Lauritzen, S.L., Spiegelhalter, D.J.: *Probabilistic Networks and Expert Systems*. Springer, New York (1999)
12. Murphy, K.P.: *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California, Berkeley, Berkeley (2002)
13. Funiak, S., Guestrin, C., Paskin, M., Sukthankar, R.: Distributed localization of networked cameras. In: *Proc. of Fifth International Conference on Information Processing in Sensor Networks (IPSN 2006)*, pp. 34–42 (2006)
14. Funiak, S., Guestrin, C., Paskin, M., Sukthankar, R.: Distributed inference in dynamical systems. In: *Twentieth Annual Conference on Neural Information Processing Systems* (2006)
15. Dellaert, F., Krauthausen, P.: A multifrontal qr factorization approach to distributed inference applied to multi-robot localization and mapping. In: *Proceedings of the American Association for Artificial Intelligence*, pp. 1261–1266 (2005)
16. Lauritzen, S.L., Spiegelhalter, D.J.: Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society, Series B* 50, 157–224 (1988)



# Optimized Particles for 3-D Tracking

Huiying Chen<sup>1</sup> and Youfu Li<sup>2</sup>

<sup>1</sup> Department of Industrial and Systems Engineering,  
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong  
mfhychen@inet.polyu.edu.hk

<sup>2</sup> Department of Manufacturing Engineering and Engineering Management,  
City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong  
meyfli@cityu.edu.hk

**Abstract.** 3-D visual tracking is useful for many of its applications. In this paper, we propose two different ways for different system configurations to optimize particle filter for enhancing 3-D tracking performances. On one hand, a new data fusion method is proposed to obtain the optimal importance density function for active vision systems. On the other hand, we develop a method for reconfigurable vision systems to maximize the effective sampling size in particle filter, which consequentially helps to solve the degeneracy problem and minimize the tracking error.

**Keywords:** 3-D tracking, particle filter, importance density function, effective sampling size.

## 1 Introduction

There-dimensional (3-D) tracking deals with continuous 3-D state estimation and update of moving objects [1]. The task of 3-D tracking is of paramount importance for many applications and has been considered from widely different perspectives of various theoretical backgrounds and interests. As one of the state-space estimation problems, 3-D tracking can be modeled with the aid of parametric models. However, due to varying degrees of uncertainty inherent in system modeling and complexity of system noise, visual system is often subject to elements of non-Gaussianity, non-linearity, and high dimensionality, which unfortunately, usually precludes analytic solutions. It is a strong belief that the issue of state measurement ultimately remains best handled within the framework of statistical inference. Instead of using linearization techniques, the estimation problem is solved directly with Bayesian methodology [2], [3]. However, the Bayesian paradigm involves calculation of high order integrals of the time state estimation. Thus, in the last few decades, many approximation filtering schemes, which are well-known as methods of Particle Filtering (PF), also known as Condensation or Sequential Monte Carlo methods (SMC) [4]-[7], have been developed to seek a simulation-based way to surmount the problems.

However, a general 3-D tracking problem with 6-DOF often requires thousands of particles [8], which can run foul of computational complexity and further interfere

real-time performance for tracking agile motion. Moreover, degeneracy phenomenon is a common problem with particle filters. As a result of degeneracy, all but one particle will have negligible weight after a few state transitions. Degeneracy implies the wastage of computational resources that a large effort is engaged to update particles whose contribution to the approximation to posterior states is almost zero. Doucet [9] has shown that the variance of the importance weights can only increase over time so that degeneracy is an inevitable phenomenon with general sequential importance sampling scheme. There are commonly three methods to tackle the degeneracy problem [10]: (1) brute force approach, (2) good choice of importance density, and (3) use of resampling. The brute force approach uses a large enough sampling size to cover the effect of weight degeneration. However, it is often impractical in real-time estimation system. The method of choosing the optimal importance density [9] can maximize the effective sampling size [11], [12], which is a suitable measure of degeneracy. The third method involves using the resampling process to reduce degenerate effects [4]. Although resampling has been employed a lot in generic particle filter to avoid degeneracy as one of the most popular methods, it introduces additional computation complexity and cannot help to reduce the number of particles.

In this paper, we intent to explore possible ways to optimize the particle filter for enhancing 3-D tracking performances. On one hand, a new data fusion method is proposed to obtain the optimal importance density function so that particle crowds can represent the posterior states in a much more efficient fashion. As a result, for achieving the same tracking accuracy, the number of particles used in 3-D tracking is greatly reduced. On the other hand, we develop a method for reconfigurable vision systems to maximize the effective sampling size in particle filter, which consequentially helps to solve the degeneracy problem and minimize the tracking error.

## 2 Developing the Particle Framework

At time step  $k$ , when a measurement  $\mathbf{y}_k$  becomes available, according to the Bayes' rule [2], the posterior probability function of the state vector can be calculated using the following equation

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) = \frac{p(\mathbf{x}_k | \mathbf{x}_{k-1})p(\mathbf{y}_k | \mathbf{x}_k)}{p(\mathbf{y}_k | \mathbf{y}_{1:k-1})}. \quad (1)$$

$P(\mathbf{y})$  is a normalizing constant, and (1) can be written as

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) \propto p(\mathbf{x}_k | \mathbf{x}_{k-1})p(\mathbf{y}_k | \mathbf{x}_k). \quad (2)$$

Suppose at time step  $k$  there is a set of particles,  $\{\mathbf{x}_k^i, i=1, \dots, N_s\}$  with associated weights  $\{w_k^i, i=1, \dots, N_s\}$  randomly drawn from importance sampling [9], [11], where  $N_s$  is the total number of particles. The weight of particle  $i$  can be defined as

$$w_k^i \propto w_{k-1}^i \frac{p(\mathbf{x}_k^i | \mathbf{x}_{k-1}^i)p(\mathbf{y}_k | \mathbf{x}_k^i)}{q(\mathbf{x}_k^i | \mathbf{x}_{k-1}^i, \mathbf{y}_{1:k})}, \quad (3)$$

where  $q(\mathbf{x}_k^i | \mathbf{x}_{k-1}^i, \mathbf{y}_{1:k})$  is the importance density function. In this paper, we use the transition prior  $p(\mathbf{x}_k | \mathbf{x}_{k-1})$  as the importance density function. Then (3) can be simplified as

$$w_k^i \propto w_{k-1}^i p(\mathbf{y}_k | \mathbf{x}_k^i). \tag{4}$$

Furthermore, if we use Grenander’s factored sampling algorithm [12], (4) can be modified as

$$w_k^i = p(\mathbf{y}_k | \mathbf{x}_k^i). \tag{5}$$

The particle weights then can be normalized using

$$w_k^{*i} = \frac{w_k^i}{\sum_{i=1}^{N_s} w_k^i} \tag{6}$$

to give a weighted approximation of the posterior density in the following form

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) \approx \sum_{i=1}^{N_s} w_k^{*i} \delta(\mathbf{x}_k - \mathbf{x}_k^i), \tag{7}$$

where  $\delta$  is the Dirac’s delta function.

### 3 Data Fusion for Importance Density Optimization

#### 3.1 Methodology

The 3D tracking task here is performed with an *active vision system* [13] using pattern projection, which is similar to a passive stereo vision system with one of the cameras replaced by a projector. Using a color-encoded structured light pattern [14], the active vision system can yield good results in 3D visual sensing with a single view.

In order to obtain better expression of posterior states, the importance density function should be moved towards the region of high likelihood. Notwithstanding, because the support valuables are different, likelihood functions cannot be used directly to modify the importance density. To surmount this problem, a pseudo likelihood function is first generated with the latest passive sensing data. Then the pseudo likelihood is projected to the importance density space and the importance density is modified by fusing the sensing data in it. The pseudo likelihood function is generated with the most current observation of certain reference feature points through passive sensing. It is a subset of the likelihood function and it can represent the likelihood function to certain extent. The advantage of the pseudo likelihood function is that it can be projected to the importance density space easily by using the inverse procedure of passive sensing observation model (a monocular camera model). The basic idea of the proposed approach is illustrated in Fig. 1. Suppose in a set of reference points, the observation of passive sensing (of a monocular camera) can be expressed as a function of the current state with noise

$$y_{k_j}^R = g_j(x_k, \zeta), \quad j = 1, \dots, N_R \tag{8}$$

where  $\zeta$  is noise and  $N_R$  is the number of reference points.

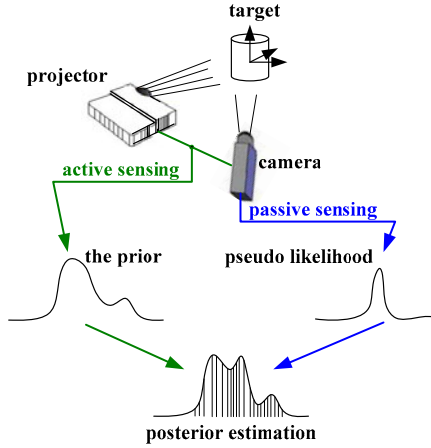


Fig. 1. Data fusion with updated passive sensing data

Equation (8) can be looked on as a pseudo likelihood function

$$y_k^R \sim \hat{L}_y = P(\mathbf{y} | x_k^R). \tag{9}$$

On the contrary, the current state can be estimated using the inverse function of (8) as

$$x_k = g^{-1}(x_{k-1}, y_k^R, \zeta), \tag{10}$$

which is in fact a projection of the pseudo likelihood to the importance density ( $x_k$ ) space

$$\hat{L}_x = P(x_k | x_{k-1}, y_k^R, \zeta). \tag{11}$$

Then (11) can be used to achieve the algorithm for data fusion as shown in Table 1.

### 3.2 Simulation Results

Fig. 2 shows the estimation error for 3D location tracking by different methods. The generic PF, which employed 800 particles, performed the best, while the extended Kalman Filter (EKF), with 100 particles, performed the worst because of its disadvantage in dealing with multi-modality. The proposed PF with data fusion, even only with 100 particles, achieved performance approximately as good as the generic PF.

With a well-expressed importance density, the proposed PF can achieve better real time performance with expedition. Simulation results demonstrate the superiority of the proposed method in comparisons with EKF and generic PF (GPF). Because the EKF does not involve calculations of sampling, it can achieve faster real-time performance with an average 0.0146s for each state. The algorithm running times for GPF with 200, 400 and 800 particles are 0.0271s, 0.0522s and 0.0998s respectively. With only 100 particles, the proposed PF excels the GPFs in running time and only spent 0.0150s in average.

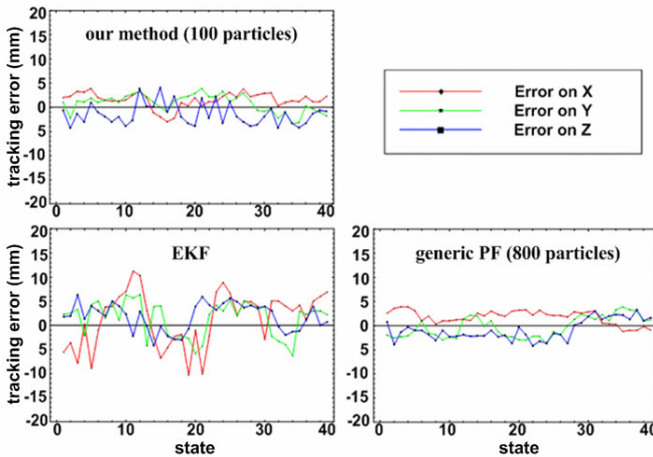
**Table 1.** Passive data fusion algorithm in particle filter

---

Assume that at the previous state  $k-1$ , we have the particle crowd  $\{x_{k-1}^i, W_{k-1}^i\}_{i=1}^N$ , then proceed as following at time  $k$

- 1. Sampling:** simulate  $x_k^i \sim P(x_k|x_{k-1})$
- 2. Calculate weights:** compute the weights according to likelihood function and conduct normalization
- 3. Pseudo likelihood computation:**  
calculate  $\hat{L}_x = P(x_k|x_{k-1}, y_k^R, \zeta)$
- 4. Data fusion:** simulate  $x_k^i$ , draw  $\alpha N$  samples from the prior  $P(x_k|x_{k-1})$  and  $(1-\alpha)N$  samples from the pseudo likelihood projection  $P(x_k|x_{k-1}, y_k^R, \zeta)$ , where  $\alpha$  is a data fusion factor,  $0 \leq \alpha \leq 1$
- 5. Update weights:** compute the weights according to the new likelihood function and conduct normalization
- 6. Resampling**

---



**Fig. 2.** Tracking accuracy comparison

### 3.3 Experimental Results

The proposed tracking method was tested with an active vision system which consists of a PULNIX TMC-9700 CCD camera and a PLUS V131 DLP projector (as shown in Fig. 3(a)). When the system is used in a visual tracking task, the projector projects a color-encoded structured light (see Fig. 3(b), [15]) onto the surface of the target object. Via triangulation, the system returns a time sequence of 3-D object positions and orientations. This provides the measurement (given in Section 2) for the tracking formulation.

We used a concave object as the target (Fig. 4), which was moved arbitrarily by hand in 3-D space to give motions with 3-DOF translational and 2-DOF rotational.

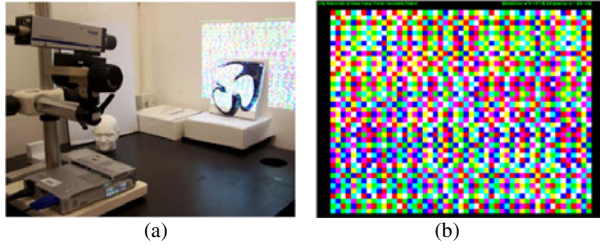


Fig. 3. The active vision system using color-encoded structured light

The tracker (formulated in Section 2 and 3) was used to estimate the target’s 5-DOF position. Here, a PF with 100 particles was employed. Since the object was moved randomly and the tracking was performed in real-time, quantitative results on tracking accuracy were hard to obtain due to lack of the ground truth. We thus re-projected the estimated object positions and orientations onto the camera image (the red circles shown in Fig. 4) for qualitative evaluation. Some examples of snap shots in the tracking are shown in Fig. 4. With a sampling rate of about 12fps, correct and reliable tracking were observed in the implementation. In Fig. 4, the tracking errors were mainly caused by the sensing itself, rather than the tracker. For example, in the frame shown in the bottom-right, relatively larger tracking error is observed. This is because the target happened to move to a position where the structured light pattern could not be detected clearly.

## 4 Dynamic View Planning for Maximizing the Number of Effective Particles

### 4.1 Methodology

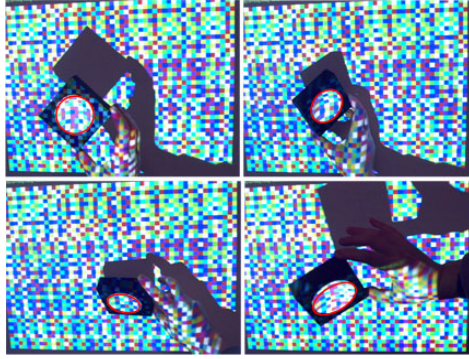
The 3-D tracking task here is performed with a *reconfigurable vision system* [16]. In this section, we intend to use the re-configurability of the vision system to reduce those effects of degeneracy in particle filter. According to [11], [12], the effective sampling size  $N_k^{eff}$  at state  $k$  is defined as

$$N_k^{eff} = \frac{N_s}{1 + \text{Var}(w_k^i)}, \tag{12}$$

where  $w_k^i$  is referred to as the “true weight” indicated in (4), and  $N_s$  is the number of samples. As  $N_k^{eff}$  cannot be evaluated exactly [10], an estimate  $\hat{N}_k^{eff}$  of  $N_k^{eff}$  can be calculated by

$$\hat{N}_k^{eff} = \frac{1}{\sum_{i=1}^{N_s} (w_k^{*i})^2}, \tag{13}$$

where  $w_k^{*i}$  is the normalized weight indicated in (12).



**Fig. 4.** Tracking a concave object with the proposed method

We then define the rate of effective particles as

$$\lambda_k^{eff} = \frac{\hat{N}_k^{eff}}{N_s}. \quad (14)$$

Finally, the view planning [16] task is achieved by computing the best configuration  $\zeta_k^*$  in the viewpoint configuration space  $C_v^k$  of the camera through the following equation

$$\zeta_k^* = \arg \min_{\zeta_k} \left( \sum_{i=1}^{N_s} (w_k^{*i})^2 \right) \Big|_{\zeta_k \in C_v^k} = \arg \max_{\zeta_k \in C_v^k} \left( \hat{N}_k^{eff} \right). \quad (15)$$

## 4.2 Simulation Results

In the first simulation, we compared the effective particles (sample size)  $\hat{N}_k^{eff}$  of the proposed view planning method to the generic particle filter with a fixed viewpoint. We used 100 particles for each methods and ran the simulation for 100 times with 9 state transitions. The average number of effective particles of the two methods are plotted as shown in Fig. 5. The generic PF without view planning obtained a very low rate of effective particles (7%), while our approach maximized the rate of effective particles  $\lambda^{eff}$  at about 53% via dynamic view planning. Our algorithm realizes view planning and achieves the best configurations of the vision system by maximizing

$\hat{N}_k^{eff}$  (or minimizing  $\sum_{i=1}^{N_s} (w_k^{*i})^2$ ). In the second simulation, we checked tracking errors of different view point configurations (camera locations) to prove that the best configuration in the sense of sampling efficiency is consistent with the best configuration in the sense of minimizing tracking error. We tested view planning results with

different position parameters. As shown in Fig. 6, average values of  $\sum_{i=1}^{N_s} (w_k^{*i})^2$  and tracking errors are plotted with different configurations and estimation states. Here,

the tracking error is defined as the distance between the estimated location and its true location. 100 tests each with 100 particles were employed. 9 viewpoint locations in camera coordinates,  $x_i = 170 \times i(\text{mm})$ , when  $i=1, \dots, 9$ , were employed. These locations were chosen empirically considering both the sensitivity and the kinematics constraints of the system.

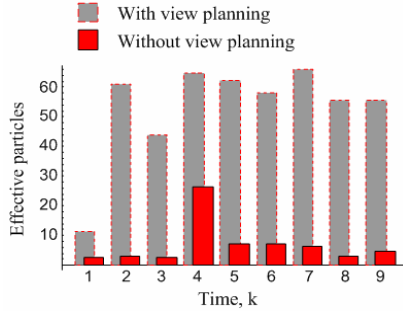


Fig. 5. Number of effective particles

It can be seen in the Fig. 6 that these two evaluation criteria shared the same tendency in viewpoint configuration. The comparison in 2-D figure at the 6<sup>th</sup> estimation state is shown in Fig. 7. Different evaluation values with their values of standard deviation are plotted. Fig. 7 clearly shows that tracking error reaches its minimum (at the 4<sup>th</sup>  $x$  location,  $x=680\text{mm}$ ) when  $\sum_{i=1}^{N_k} (w_k^{*i})^2$  reaches its minimum value. In other

words, the view planning driven by optimizing particle sampling actually minimizes the tracking error and improves tracking performance.

Then we compared our method with the “centering” view planning method and the “error-orientated” view planning method. The centering method has been usually adopted in visual servoing [17], which controls the view point to keep the image feature of the target object always at the center point of the image screen. In the “error-orientated” view planning method [18], the view planning process was directly driven by minimizing the estimated tracking error. The tracking errors of the aforementioned methods are plotted in Fig. 8, and their tracking performances are evaluated in Table 2. In these tests, our view planning method by effective particles was superior to others in tracking performance with the smallest tracking error and reasonable tracking speed. When the resampling method was used, although it could reduce the effects of weights degeneracy, the tracking error was large when using a small number of particles, whereas tracking speed was slow when using a large number of particles. Without testing the particle weights to obtain the best configuration, the centering method showed its advantage in tracking speed. However, because it can only compensate for a part of the sensing error and because it still suffers from particle degeneracy, its tracking error was larger than our method. The error-orientated method was supposed to be the one that could achieve good results in tracking accuracy. However, practically, it needs direct backward calculations and random search in the parameter space which may preclude unique solutions and yet, cannot overcome the degeneracy, which affects its tracking accuracy and the achievable tracking speed as well.



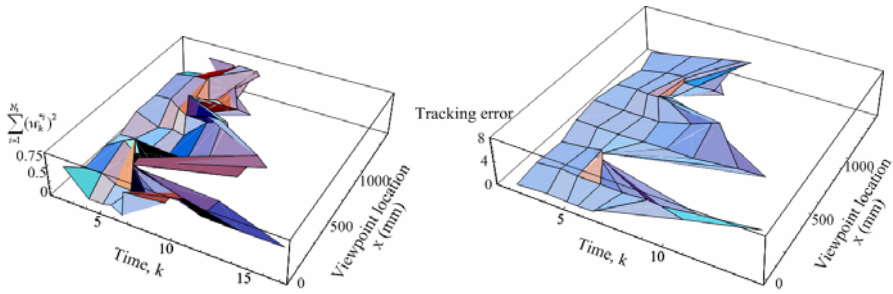


Fig. 6. Testing different evaluation criteria with position parameter  $x$  (3-D)

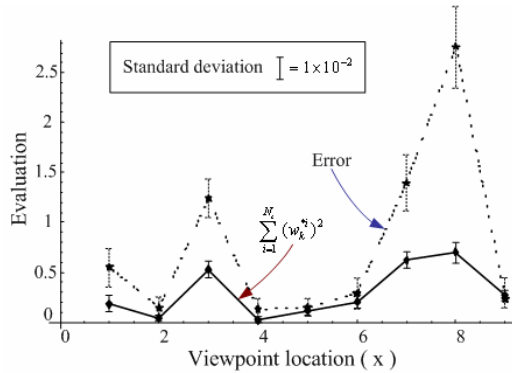


Fig. 7. Testing different evaluation criteria with position parameter  $x$  (2-D)

Our method with effective particles minimizes tracking error by revealing the system to a better swarm of importance samples and interpreting the posterior state in a better way. Furthermore, it reduces particles’ degeneracy significantly so that a relative smaller particle crowd can be used to achieve the same level of tracking performance, and thus increases possible tracking speed.

### 4.3 Experimental Results

The implementation of the proposed view planning method was conducted using our reconfigurable vision system, with a PC-based IM-PCI system and a variable scan frame grabber. This system supports many real-time processing functions including some feature extraction such as edge detection. Our algorithms were developed in VC++ programming language and run as imported functions by ITEX-CM. The system setup consists of a color CCD camera (model Pulnix TMC-6), with a resolution of  $640 \times 480$  pixels, a pan-tilt unit (model PTU-46-17.5) for 2-axis angular motion, and a linear motion system with a guideway (model KK86-20) and motion controller (model Elmo BAS-3/320-2). A photo of this 3-DOF system is shown in Fig. 9.

For simplicity’s sake, we used a point object for the experiment and made the object undergo uniform circular motion around the centre with different diameters on a plane perpendicular to the optical axis of the camera at its original location

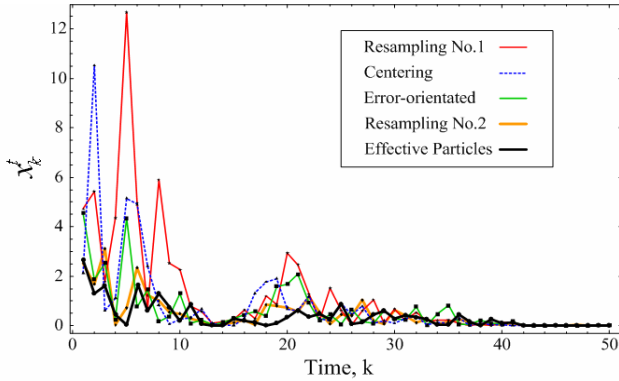


Fig. 8. Tracking errors using different methods

Table 2. Tracking performance with different methods

Evaluation Method	Total number of particles	Average relative tracking error	Average estimation time (s)	Average rate of effective particles
Effective particles	100	2.7%	0.032	49.5%
Resampling PF No. 1	100	7.5%	0.030	6.6%*
Resampling PF No. 2	1000	4.5%	0.266	7.6%*
Centering	100	6.5%	0.027	6.1%
Error-orientated	100	5.5%	0.111	7.1%

(\* This is the rate of effective particles  $\lambda^{eff}$  before resampling. After resampling, the rate of effective particles is compulsory modified to 100%. Even though it shows no improvement in tracking performance. )

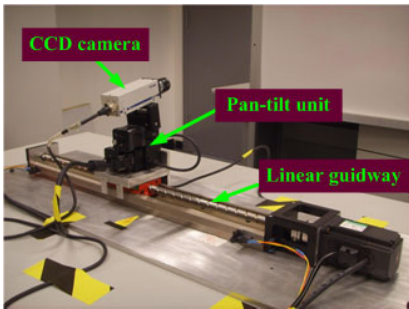


Fig. 9. Reconfigurable vision system with 3 DOFs

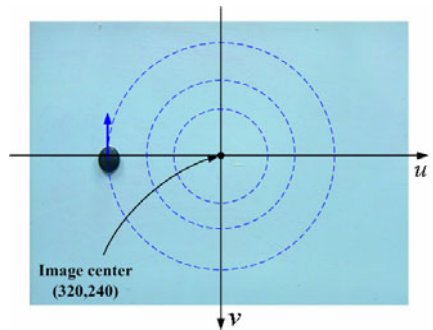


Fig. 10. Uniform circular motion at different diameters

(see Fig. 10). Using this motion, we can eliminate both influence of image feature location and influence of velocity on tracking error (see [1] for detailed discussion).

The average tracking errors on different concentric circles with different methods were calculated and are given in Table 3. These results show that our view planning method with effective particles can achieve small tracking error then other two methods.

**Table 3.** Absolute tracking errors with different view planning methods

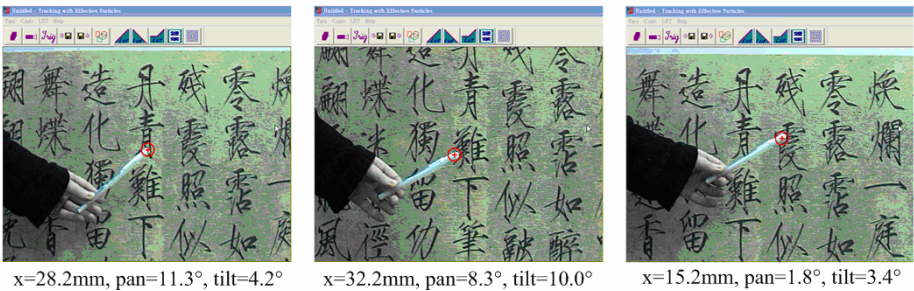
Test Diameter $2r$ (pixel)	Effective Particles		Centering		Error-Orientated	
	Mean Tracking Error (mm)	Standard Deviation ( $\text{mm}^2$ )	Mean Tracking Error (mm)	Standard Deviation ( $\text{mm}^2$ )	Mean Tracking Error (mm)	Standard Deviation ( $\text{mm}^2$ )
400	6.2	19.1	7.2	18.5	8.8	21.6
360	6.1	21.1	9.1	19.7	9.0	19.1
320	4.1	18.8	8.3	16.1	8.9	19.6
280	5.6	17.8	6.6	18.8	7.0	19.6
240	5.9	19.5	6.5	15.5	5.5	18.5
200	5.3	19.3	6.2	16.5	8.4	19.5
160	4.6	17.1	7.0	17.3	4.7	18.1
120	5.2	16.6	6.3	15.6	5.9	19.6
80	5.8	18.3	7.5	16.1	5.9	18.6
Mean	5.4	18.6	7.2	17.1	7.1	19.4

**Table 4.** Tracking speed with different view planning methods

Method	Effective Particles	Centering	Error-Orientated
Tracking Speed (fps)	24	24	7

Average tracking speeds with different methods are listed in Table 4. Because our method and centering method do not involve random search procedure that the error-orientated method uses, they both achieved very nice real-time performance.

Our method with effective particles was then implemented to track a pen tip which was moving randomly with an average speed about 1m/s. The pen tip was detected and tracking based on the segmentation with color and contour cues. Some examples of snapshots in the tracking with their corresponding viewpoint locations (best configurations) are shown in Fig. 11. In this experiment, every pair of two sequential frames was employed and compared to calculate the depth information  $z$ , and beside the current configuration information, the Chinese calligraphy background was used as correspondence between every two frames for further modification. Even with this process, a tracking rate of about 17fps was achieved in the implementation. We then re-projected the estimated 3-D locations of the pen tip onto the image space for tracking error analysis. The red target marks in Fig. 11 represent those estimations from



**Fig. 11.** Dynamic view planning in 3D tracking by our method

our tracking algorithm. Experimental results show that the tracking was conducted with very good accuracy, with an average tracking error of 2.8 pixels.

## 5 Conclusion

In this paper, two different methods for active vision and reconfigurable vision system are explored respectively to optimize particle filter for enhancing 3-D tracking performance. A new data fusion method has been proposed to obtain the optimal importance density function for active vision so that particle crowds can represent the posterior states in a much more efficient fashion. As a result, for achieving the same tracking accuracy, the number of particles used in 3-D tracking is greatly reduced. We have also developed a method for reconfigurable vision systems to maximize the effective sampling size in particle filter, which consequentially helps to solve the degeneracy problem and minimize the tracking error. Simulation and experimental results have verified the effectiveness of our methods.

## References

1. Chen, H.Y., Li, Y.F.: Enhanced 3D tracking using nonsingular constraints. *Optical Engineering* 45(10), 107202-1-12 (October 2006)
2. Bernardo, J.M., Smith, A.F.M.: *Bayesian Theory*. Wiley, Chichester (1994)
3. Zhang, Y., Ji, Q.: Active and dynamic information fusion for multisensor systems with dynamic bayesian networks. *IEEE Transactions on Systems, Man and Cybernetics (Part B)* 36(2), 467–472 (2006)
4. Gordon, N.J., Salmond, D.J., Smith, A.F.M.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: *IEE Proc. For Radar and Signal Processing*, vol. 140(2), pp. 107–113 (1993)
5. Doucet, A., Freitas, N.D., Gordon, N.: *Sequential Monte Carlo Methods in Practice*. Springer, Heidelberg (January 2001)
6. Isard, M., Blake, A.: CONDENSATION – conditional density propagation for visual tracking. *International Journal of Computer Vision (IJCV)* 29(1), 2–28 (1998)
7. Scott, W.R., et al.: Performance-orientated view planing for automatic model acquisition. In: *Proc. of the 31th Int. Symposium on Robotics*, pp. 314–319 (2000)
8. Fox, D.: KLD-sampling: adaptive particle filters. In: *Proc. of Conf. on Neural Information Processing Systems* (2001)
9. Doucet, A.: On sequential simulation-based methods for Bayesian filtering, Technical Report CUED/F-INFENG/TR 310, Dept. of Engineering, Cambridge University (1998)
10. Arulampalam, M.S., et al.: A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing* 50(2) (February 2002)
11. Bergman, N.: *Recursive Bayesian Estimation: Navigation and tracking applications*, Ph. D. Dissertation, Linkoping Univ., Linkoping, Sweden (1999)
12. Carpenter, J., Clifford, P., Fearnhead, P.: Improved particle filter for nonlinear problems. In: *IEE Proceedings on Radar, Sonar, Navig.*, vol. 146(1), pp. 2–7 (February 1999)
13. Besl, P.: Active optical range imaging sensors. *Machine Vision and Applications* 1, 127–152 (1988)
14. Griffin, P.M., Narasimhan, L.S., Yee, S.R.: Generation of uniquely encoded light patterns for range data acquisition. *Pattern Recognition* 25(6), 609–616 (1992)

15. Zhang, B., Li, Y.F., Wu, Y.: Self-recalibration of a structured light system via plane-based homography. *Pattern Recognition* 40(4), 1368–1377 (2007)
16. Chen, H.Y., Li, Y.F.: Dynamic view planning by effective particles for 3D tracking. *IEEE Transactions on Systems, Man and Cybernetics (Part B)* 39(1), 242–253 (2009)
17. Malis, E., Chaumette, F., Boudet, S.: 2-1/2-D visual servoing. *IEEE Transactions on Robotics and Automation* 15, 238–250 (1999)
18. Spletzer, J.R., Taylor, C.J.: Dynamic sensor planning and control for optimally tracking targets. *Int. Journal of Robotics Research* 22(1), 7–20 (2003)

# Design of a Finger-Tip Flexible Tactile Sensor for an Anthropomorphic Robot Hand

Yuan-Fei Zhang<sup>1</sup>, Yi-Wei Liu<sup>1</sup>, Ming-He Jin<sup>1</sup>, and Hong Liu<sup>1,2</sup>

<sup>1</sup> State Key Laboratory of Robotics and System, Harbin Institute of Technology,  
150001 Harbin, China

<sup>2</sup> Institute of Robotics and Mechatronics, German Aerospace Center,  
82230 Munich, Germany  
zyf\_hit@yahoo.com.cn

**Abstract.** In this study, a thin and flexible tactile sensor using pressure-conductive rubber for a HIT/DLR Hand II dextrous robot hand is developed. A method of the curved surface approximate development is used to design the sensor shape, ensuring that the sensor can fine cover the three-dimensional (3D) finger-tip surface. In order to weaken crosstalk current and simplify electrical circuit of the sensor, a sampling-voltage-feedback-non-target-scanned-sampling-electrode (SVFNTSSE) method is proposed. Then a structure of tactile signal acquisition and processing circuit based on this method is described. Finally, the experiment of weakening crosstalk based on the SVFNTSSE method verifies the validity of this method, and the results of the other experiment show that this sensor system has integrated into the dextrous robot hand.

**Keywords:** Tactile sensor; Pressure-conductive rubber; Anthropomorphic robot hand; Crosstalk.

## 1 Introduction

Up until now, the main transduction methods of tactile sensors used in dextrous robot hands are optical [1], [2] and resistive [3], [4], [5], [6]. Along with the development of dextrous robot hands towards the miniaturized, highly integrated and anthropomorphic direction, it is necessary to attach tactile sensors only on the surface of dextrous robot hands. Therefore, the tactile sensor to be applied in an anthropomorphic dextrous robot hand, which has a curved surface, should be thin and flexible. The optical tactile sensor is not suitable for an anthropomorphic dextrous robot hand, for difficult to be thin and flexible. Hence, all optical tactile sensors, which have been used in dextrous robot hands, presented in literature are still in the experimental prototype stage. However, the resistive tactile sensor may be thin and flexible. Consequently, in literature, there are more applications of resistive tactile sensors than optical tactile sensors in dextrous robot hands.

Raparelli *et al.* used Force Sensing Resistor (FSR), manufactured by Interlink Company, as the tactile sensor of a pneumatic anthropomorphic hand [5]. The sensing

element of FSR is printed on a thin film. The film is a pliable sheet, but it is not elastic. Therefore the film can only cover a three-dimensional (3D) developable surface, but it cannot exactly cover a 3D undevelopable surface. Kawasaki *et al.* developed a distributed tactile sensor for Gifu Hand II [3]. The distributed tactile sensor, which has grid pattern electrodes and uses conductive ink, is mounted on the hand surface. And this sensor can cover the most area of Gifu hand II and increase tactile sensing area. However, as the electrical wires of the sensor are outside of dextrous robot hand, it will affect the operating safety of the hand in the unstructured environment. Shimojo *et al.* designed a thin and flexible tactile sensor using pressure-conductive rubber with stitched electrical wires [4]. The sensor can cover 3D entities, currently only used for contact/non-contact state detection. Although the sensor using this method can be effective to improve the surface strength to resist shear force, this method makes the electrodes of the sensor on the double sides of the sensitive sheet. This is unfavorable, since pressure-conductive rubber is usually flexible, whereby the outer-side electrode is exposed to a bending stress reducing the lifetime of the sensor. Cannata and Maggiali developed a new fully embedded tactile sensor system [6]. The tactile sensor has been designed to be installed on a dextrous robot gripper (MAC-HAND) and consists of a matrix of 64 electrodes, etched on a flexible PCB (Printed Circuit Board) covered by a conductive rubber layer. The layout of the electrodes is on the same side, avoiding the disadvantage when a lay of double-side electrodes is used. However, the circuits of the tactile signal processing hardware and the entity of the sensor are sharing the same flexible PCB, resulting in the whole system having to be replaced even a partial failure occurred.

Through analyzing the characteristics of the past resistive tactile sensors applied in dextrous robot hands and the highly integrated features of a HIT/DLR Hand II dextrous robot hand [7], we develop a resistive tactile sensor that is thin and flexible and can be effective to cover the curved finger-tip surface of the HIT/DLR Hand II dextrous robot hand. In this paper, firstly, the design of the sensor is introduced. Secondly, the tactile signal acquisition and processing system and the communication system are described. Finally, some experimental results are given.

## 2 Design of the Tactile Sensor

The HIT/DLR Hand II dextrous robot hand is a multisensory and integrated five-fingered hand with in total 15 degree of freedoms (DOFs). To achieve a high degree of modularity, all five fingers are identical. Each finger has 3 DOFs and four joints, the last two joints are coupled by means of a “∞” figure wire mechanism. All actuators are integrated in the finger directly, the electronics and communication controllers are fully integrated in the finger in order to realize modularity of the hand and minimize weight and amount of cables needed for a hand. See [7] for more details of this robot hand. Fig. 1 shows a finger of the HIT/DLR Hand II dextrous robot hand with the developed tactile sensor. This sensor is thin and flexible, mounted on the 3D surface of the finger-tip face. The following are about to describe the sensor principle, the sensor structure and the process of the sensor shape design in detail.

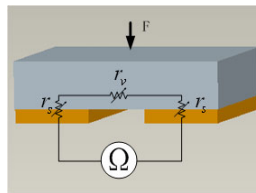


**Fig. 1.** Finger of the HIT/DLR Hand II dextrous robot hand with the finger-tip flexible tactile sensor

## 2.1 Principle of the Sensor

The principle of the tactile sensor is based on the resistive mechanism. This sensor adopts the pressure-conductive rubber CSA (the product of Yokohama Rubber Co., Ltd.) as the pressure sensitive material. The rubber is a composite of silicon rubber and graphite particles, and is available in gray-black flexible sheet form, 0.5 mm in thickness. It has many excellent features: such as elasticity, elongation, flexibility, and durability [8].

The surface resistive effect [9] is mainly applied to increase the sensitive of the tactile sensor. As shown in Fig. 2, the volume resistance  $r_v$  of the pressure sensitive material between the electrodes and the surface resistance  $r_s$  are varying with the applied load. However, the value of  $r_s$  is varying more observably than the value of  $r_v$  [10], when the pressure  $F$  is changed. The resistance  $r$  of the tactile cell, decreased with the increase of the load, is equal to two times  $r_s$  plus  $r_v$ . And the information of contact would be obtained by measuring the value of  $r$ .

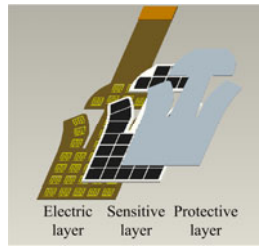


**Fig. 2.** Schematic of the surface resistive effect

## 2.2 Structure of the Sensor

The layouts of the tactile sensor electrodes presented in literature are divided into two types: a double-side electrode layout and a single-side electrode layout. Since the pressure-conductive rubber is flexible, using the layout of double-side electrodes is unfavorable. There must be one side electrode exposed to a bending stress reducing the lifetime of the sensor, when pressure is applied repeatedly. So the layout of single-side electrodes is adopted by the developed sensor. The sensor has three layers: an electric layer, a sensitive layer and a protective layer, as shown in Fig. 3. The thickness of this sensor and the size of a tactile cell are 0.65 mm and 3 mm × 3 mm, respectively.





**Fig. 3.** Structure of the finger-tip flexible tactile sensor

The electric layer is made of a double-layer flexible PCB and mounted on the finger-tip face of the HIT/DLR Hand II dextrous hand by special glue. The flexible PCB, in the size of  $50 \text{ mm} \times 19.6 \text{ mm} \times 0.1 \text{ mm}$  (length  $\times$  width  $\times$  thickness), comprises 36 pairs of gold-plated comb-shaped electrodes, the size of which is  $2.1 \text{ mm} \times 2.1 \text{ mm}$ . By means of arraying the electrodes of the sensor, there only need 13 out pins to connect with a tactile signal acquisition and processing circuit.

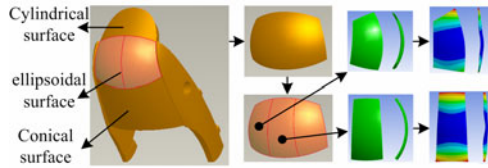
The sensitive layer comprises the pressure-conductive rubber and insulating silicon-rubber glue. The rubber sheet is divided into a small part for each tactile element, embedded into an insulating meshwork. Here the insulating silicon-rubber glue is applied not only to eliminate the internal interference of the sensitive material between adjacent sampling electrodes and adjacent driving electrodes, but also to glue the sensitive material on the electric layer (be sure that the glue does not exist between the sensitive material and the electrodes of the electric layer).

The protective layer is an insulating silicon-rubber film. And the effects of this layer are not only to protect the sensitive layer, but also to increase the friction coefficient of the sensor surface and to vary the measurement scale of the sensor in a certain extent.

### 2.3 Shape of the Sensor

In order to make the appearance of HIT/DLR Hand II dextrous robot hand like the human hand, the finger-tip surface has been designed to a 3D surface. From top to bottom, a cylindrical surface, an ellipsoidal surface and a conical surface compose this 3D surface with differentiable patch joins, as show in Fig. 4. Because of restrictions of the flexible PCB processing technology, only planar flexible PCB can be produced. Therefore, the shape design of the flexible PCB becomes a key research content in order to make the sensor fine cover the finger-tip surface. To simplify the issue, the research of surface fitting is reversed into the study of surface flattening. Because the cylindrical surface and the conical surface are developable surfaces except the ellipsoidal surface, the focus issue becomes how to make a reasonable approximate development of the ellipsoidal surface.

There are three steps to analyze the approximate development of the ellipsoidal surface. Firstly, extract the ellipsoidal surface from the finger-tip face; secondly, insert darts in the extracted surface to split it into several sub-surfaces; finally, approximately develop the sub-surfaces. Though the more darts are inserted in the surface, the higher development precision is acquired, this will result in the layout design of the flexible



**Fig. 4.** Analyzing process of the approximate development of the ellipsoidal

PCB more difficult. Therefore, the number of the darts should be minimized to meet the approximate development precision requirement. In this study, two darts are inserted in the ellipsoidal surface. And the surface is divided into three sub-surfaces (trisect the upper and lower lines, respectively, and then use two specific planes to split this surface). The shapes of the sub-surface contours can be obtained through the software of ANSYS 11.0 analysis, as shown in Fig. 4. This provides the guidance for the approximate development of the sub-surfaces. So, the inside edges of two side-split sub-surfaces are instead of arc curves, and the other remaining edges are instead of straight lines. Consequently, the two-dimensional (2D) pattern of the approximate development of the ellipsoidal surface is shown in Fig. 5(a).

Next we are about to analyze the development precision of the ellipsoidal surface from the aspect of the overall error that includes the relative angle error  $E_\theta$  of two grid lines, the relative length error  $E_L$  of the grid line and the relative area error  $E_S$  of the grid.

As the ellipsoidal surface has a surface symmetry property before and after the approximate development, there only need to analyze the relative error of the parameters  $\theta_1$ - $\theta_6$ ,  $L_1$ - $L_7$  and  $S_1$ - $S_2$ . The results that all kinds of errors are less than 3% are shown in Table 1, Table 2 and Table 3. Hence, this 2D surface can well fit the ellipsoidal surface.

**Table 1.** The relative angle errors

	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$	$\theta_6$
3D model (°)	120.60	86.23	90.81	101.90	93.77	89.19
2D model (°)	121.60	85.76	92.18	99.23	92.59	87.41
$E_\theta$ (%)	0.83	1.77	1.51	-2.62	-1.26	-2.00

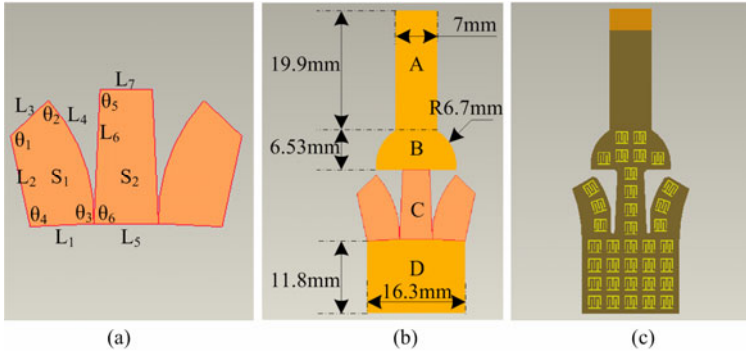
**Table 2.** The relative length errors

	L1	L2	L3	L4	L5	L6	L7
3D model (mm)	5.48	8.00	4.42	11.18	5.48	11.18	4.42
2D model (mm)	5.42	7.98	4.44	11.50	5.48	11.51	4.44
$E_L$ (%)	1.09	0.25	0.45	2.86	0	2.95	0.45

**Table 3.** The relative area errors

	S1	S2
3D model (mm)	52.77	57.75
2D model (mm)	51.55	57.04
$E_S$ (%)	-2.31	1.23

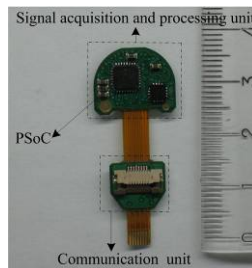
According to the 2D pattern of the ellipsoidal surface flattening, a 2D shape of the sensor is designed, as shown in Fig. 5(b). The area of A is the electrical alignment zone of the tactile sensor; the areas of B, C and D compose the tactile perception zone, and are used to cover the cylindrical surface, the ellipsoidal surface and the conical surface of the finger-tip face, respectively. Given the current restrictions of the flexible PCB processing technology, a 2D shape of the flexible PCB is designed, as shown in Fig. 5(c). Finally, the actual fitting experiment verifies that the designed shape of the sensor is fine.



**Fig. 5.** (a) 2D pattern of the approximate development of the ellipsoidal surface. (b) Shape of the tactile sensor. (c) Shape of the flexible PCB.

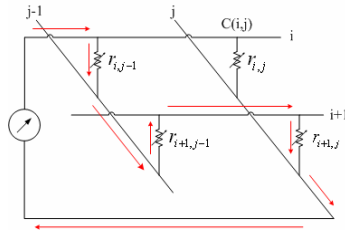
### 3 Design of the Electrical System

A Programmable System-on-Chip (PSoC), which integrates a large number of digital and analog modules that can be configured some essential functional device (such as ADC and DAC), is applied as the controller. Using this kind chip can be effective to reduce the external support circuitry complexity and improve the integration of the tactile signal acquisition and processing circuit. The entity of the circuit, which can be embedded in the fingertip (the size of approximately available space is  $16\text{ mm} \times 13\text{ mm} \times 9\text{ mm}$ ) of the robot hand, is shown in Fig. 6. According to the different functions of the circuit parts, the circuit can be divided into two functional units: a signal acquisition and processing unit and a communication unit.



**Fig. 6.** Tactile signal acquisition and processing circuit



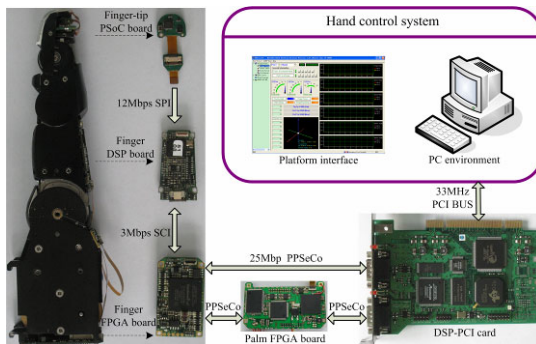


**Fig. 8.** Schematic of the crosstalk between adjacent elements

### 3.2 Communication System

A flowing path of the digital signal of tactile sensor is shown in Fig. 9. First, the tactile sensor data, packaged by the Finger-tip PSoC board, are sent to the Finger DSP (Digital Signal Processing) board via the SPI (Serial Peripheral Interface) bus; next the Finger DSP board transmits a data package, which consists of the tactile sensor data, the joint torque data and the joint angle data, to the Finger FPGA (Field-programmable Gate Array) board through the SPI serial bus; and then the Finger FPGA board sends a data package of the entire finger directly to the DSP-PCI card, or indirectly through the Palm FPGA board to the DSP-PCI card by the point-to-point serial communication (PPSeCo) bus; at last the DSP-PCI card uploads the received data to PC.

In order to diminish the communication time between the Finger-tip PSoC board and the Finger DSP board, the operating frequency of the SPI module structured in the PSoC should be set to the highest frequency (12 MHz). However, because of the clock frequency constraint of the controller, the cost time of read/write an 8 bit register is 12.6  $\mu$ s that is significantly larger than 666.7 ns which is the cost time of receive/send one byte. Therefore, to make the SPI communication working normally, the SPI module of the Finger-tip PSoC board must be set to the master mode, while the SPI module of the Finger DSP board is set to slave mode. And the final operating frequency of the tactile sensor system is 107.3 Hz via considering the time of the tactile signal acquisition and the SPI communication.



**Fig. 9.** Communication system

## 4 Experiment and Discussions

### 4.1 Experiment of Weakening Crosstalk Based on the SVFNTSSE Method

In order to verify the validity of the SVFNTSSE method, four adjacent tactile cells, which are Cell-06, Cell-07, Cell-11 and Cell-12, are selected as the experimental objects, as shown in Fig. 10. And an external force is only applied on the elements of Cell-07, Cell-11 and Cell-12, and then the output voltage of the each tactile cell can be observed. When not using the SVFNTSSE method, the output voltage of Cell-06 is seriously interfered (as shown in Fig. 10(a), the output voltage of Cell-06 is approximately equal to the others' at 8 second); but when using this method, the output voltage of Cell-06 is almost not affected, as shown in Fig. 10(b). Therefore, the experimental results verify the validity of the method.

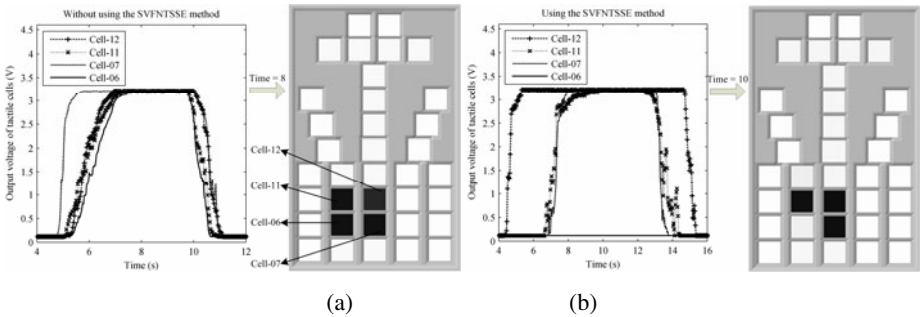


Fig. 10. Crosstalk analysis

### 4.2 Experiment of Studying Some Sensor Characteristics

This tactile sensor has nonlinear characteristics such as hysteresis and creep. These characteristics have been studied via a measurement system, as shown in Fig. 11. This system consists of a loading platform and a graph visualization system. The measurement processes are as follows: first, apply the discrete load on the sensor by the platform through the loading line, and then transmit the signal of the sensor to the PC via serial communication bus, the final process the received data by the graph visualization system. In order to improve the accuracy of the measurement, four adjacent tactile elements were selected, the pressure applied on each element should be equal, and then the average value of the element outputs should be taken as the output of the sensor. To ensure the balance of the pressure, two measures were taken: design a loading head with a flexible layer and regulate the angle of the loading head around its axis until the output value of every pressured element was approximately equal to each other. After achieving the necessary operation, we continuously loaded the discrete load and then continuously unloaded the discrete load on the sensor, two times. In this process, the sensor outputs that were average outputs of the pressured units

were acquired. The hysteresis characteristic curve and creep characteristic curve of the sensor are shown in Fig. 12(a) and Fig. 12(b), respectively. Fig. 12(a) also shows that the pressure measurement range of the sensor is about 0-600 kPa.

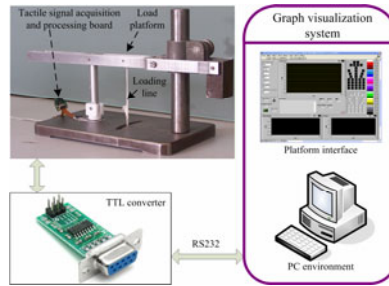


Fig. 11. Measurement system of the tactile sensor

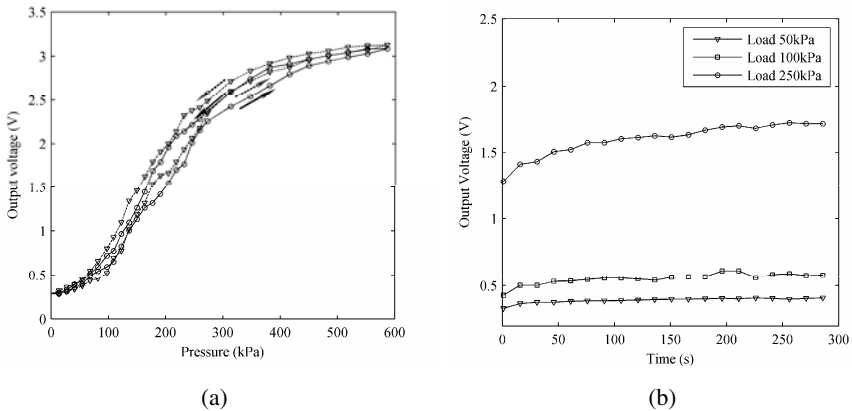
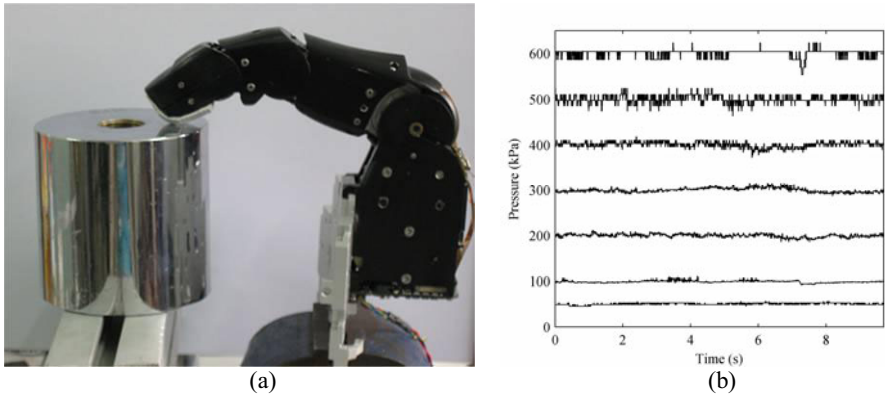


Fig. 12. (a) Hysteresis characteristic curve of the tactile sensor. (b) Creep characteristic curve of the tactile sensor.

### 4.3 Experiment of the Contact Force Control Based on the Tactile Sensor

After studying the nonlinear characteristics of the sensor, we installed the sensor on the finger-tip surface of a finger of the HIT/DLR Hand II dextrous robot hand, to build the experimental hardware system, as shown in Fig. 13(a). In this experiment, a “PI + friction compensation” control method was used, such nonlinear characteristics were taken into account for signal processing, and the desire contact force were set to 50 kPa, 100 kPa, 200 kPa, 300 kPa, 400 kPa, 500 kPa and 600 kPa, respectively. The experimental results (as shown in Fig. 13(b)) illustrate that the whole control system can maintain constant contact force based on the feedback signal of the tactile sensor. Therefore, the signal of this sensor would be applied in grasping force control of the multifingered hand in the future.



**Fig. 13.** (a) Platform for the contact control experiment. (b) Results of the constant contact force control based on the tactile.

## 5 Conclusion

In this paper, a thin and flexible resistive tactile sensor has been presented, and the embedded electrical system for tactile signal acquiring and processing has been described. The shape of the sensor is designed via using 3D surface approximate development method, ensuring that the sensor fine fits the finger-tip surface of the HIT/DLR Hand II dextrous robot hand. And the proposed SVFNTSSE method not only can be effective to weaken the crosstalk current, but also can simplify the circuit structure.

For the tactile sensor system, the effective measuring range is approximate 0-600 kPa, and the resolution and the operating frequency are 8 bits and 107.3 Hz, respectively. The contact force control experiment shows that the tactile sensor system has been integrated into the control system of the HIT/DLR Hand II dextrous robot hand, which establishes a good foundation for the tactile information used in the multifingered coordination in the future.

## Acknowledgements

This project is supported by the National High Technology Research and Development Program of China (863 Program) (No. 2008AA04Z203) and Self-Planned Task (No. SKLR200801A01) of State Key Laboratory of Robotics and System (Harbin Institute of Technology).

## References

1. Jeong, D.-H., Chu, J.-U., Lee, Y.-J.: Development of KNU hand with infrared LED-based tactile fingertip sensor. In: International Conference on Control, Automation and Systems, COEX, Seoul, Korea, pp. 1156–1161 (2008)



2. Ueda, J., Ishida, Y., Kondo, M., Ogasawara, T.: Development of the NAIST-hand with vision-based tactile fingertip sensor, Barcelona, Spain, pp. 2332–2337 (2005)
3. Kawasaki, H., Komatsu, T., Uchiyama, K.: Dexterous anthropomorphic robot hand with distributed tactile sensor: Gifu hand II. *IEEE/ASME Transactions on Mechatronics* 7, 296–303 (2002)
4. Shimojo, M., Namiki, A., Ishikawa, M., Makino, R., Mabuchi, K.: A tactile sensor sheet using pressure conductive rubber with electrical-wires stitched method. *IEEE Sensors Journal* 4, 589–596 (2004)
5. Raparelli, T., Mattiazzo, G., Mauro, S., Velardocchia, M.: Design and development of a pneumatic anthropomorphic hand. *Journal of Robotic Systems* 17, 1–15 (2000)
6. Cannata, G., Maggiali, M.: An embedded tactile and force sensor for robotic manipulation and grasping. In: 5th IEEE/RAS International Conference on Humanoid Robots, Tsukuba, Japan, pp. 80–85 (2005)
7. Fan, S.-W., Liu, Y.-W., Jin, M.-H., Lan, T., Chen, Z.-P., Liu, H., Zhao, D.-W.: Research on the mechanics of the HIT/DLR Hand II anthropomorphic five-finger dexterous hand. *Harbin Gongcheng Daxue Xuebao/Journal of Harbin Engineering University* 30, 171–177 (2009)
8. Pressure sensitive electric conductive elastomer, <http://www.scn-net.ne.jp/eagle/CSAEnglish.html>
9. Weiss, K., Worn, H.: The working principle of resistive tactile sensor cells. In: Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States, pp. 471–476 (2005)
10. Ming, X., Huang, Y., Xiang, B., Ge, Y.: Design of flexible three-dimensional force tactile sensor. *Huazhong Keji Daxue Xuebao (Ziran Kexue Ban)/Journal of Huazhong University of Science and Technology (Natural Science Edition)* 36, 137–141 (2008)
11. Tise, B.: A compact high resolution piezoresistive digital tactile sensor. In: IEEE International Conference on Robotics and Automation, New York, NY, USA, pp. 760–764 (1988)
12. Purbrick, J.A.: A Force Transducer Employing Conductive Silicone Rubber. In: First International Conference on Robot Vision and Sensory Controls, Stratford-upon-Avon, United Kingdom, pp. 73–80 (1981)
13. Hillis, W.D.: A high-resolution imaging touch sensor. *International Journal of Robotics Research* 1, 33–44 (1982)
14. Liu, Y.W., Jin, M.H., Wei, R., Cai, H.G., Liu, H., Seitz, N., Gruber, R., Hirzinger, G.: Embedded FPGA-based control of the HIT/DLR hand. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Monterey, CA, United States, pp. 582–587 (2005)

# Author Index

- Abdelgawad, Amr I-288  
Abdul-Rahman, Hussein S. II-717  
Adachi, Masaru I-226  
Alaqtash, Murad I-288  
Al-Bluwi, Ibrahim II-116  
Alboul, Lyuba II-717  
Aljanaideh, Omar I-104  
Alpen, Mirco II-215  
Antoun, Shérine M. II-130
- Bai, Yucheng II-227  
Bi, Qingzhen II-553  
BoYang, I-530  
Brower, Richard I-288  
Bui, Quoc Phuong I-442
- Cabibihan, John-John II-498, II-683  
Cao, Heng I-590  
Cao, Jiangtao II-486  
Cao, Qixin I-226  
Cao, Zhiqiang I-538  
Chang, Yau-Zen II-154  
Che, Demeng I-47  
Chen, Bing II-22  
Chen, Daguo II-475  
Chen, Genliang II-252  
Chen, Hui II-1  
Chen, Huiying I-749  
Chen, Peihua I-226  
Chen, Qiang II-56, II-695  
Chen, S.Y. I-300  
Chen, Wei II-754  
Chen, Xiang I-692  
Chen, Xiao I-590  
Chen, Xinpu I-81  
Chen, Xubing II-345  
Chen, Yixiong II-417  
Chen, Zhaopeng I-1  
Chen, Zheng I-159  
Cheng, Chen I-499  
Cheng, Shili I-409  
Chin, Chungwei II-683  
Choi, Jong-suk I-582
- Chou, Yu-Shin II-154  
Chuang, Chin-Wen I-266
- Dai, Hongjun II-695  
Dai, Jian S. I-430  
Dai, Lei I-171  
Dario, Paolo I-477  
Deng, Hua II-329, II-339  
Deng, Lei I-548, II-191  
Deng, Mingcong I-135  
Dijk, Judith I-665  
Dillmann, Ruediger I-26  
Ding, Han II-553, II-576  
Ding, Weili II-368  
Ding, Ye II-541  
Dong, Wei I-600  
Dorst, Leo I-665  
Du, Jian II-285  
Du, Juan I-92  
Du, Yingkui I-609  
Du, Zhengchun II-276
- Elnagar, Ashraf II-116  
Esteban, Isaac I-665  
Ewert, Daniel II-44
- Fan, Baojie I-609  
Fan, Hongchao II-390  
Fan, Huijie I-600  
Fan, Shaowei I-1, I-38  
Fang, Yinfeng I-300  
Fang, Zhou II-191  
Feng, Ying I-92  
Fu, Ruiqing I-548  
Fu, Xin I-477, II-296  
Fu, Xingwei II-227  
Fu, Zhuang I-400, I-519
- Ganapathy, Velappa II-624, II-636  
Gao, Zhen I-418  
Gasteratos, Antonios II-705  
Ge, Shuzhi Sam II-498, II-683  
Gong, Guofang I-329  
Gong, Hu II-67

- Goto, Takafumi II-97  
 Gu, Dongbing I-724  
 Gu, Guoying I-114  
 Guan, Enguang I-400, I-519  
 Guo, Qiang II-565  
 Guo, Qiuquan II-649  
 Guo, Wei I-257  
 Guo, Xinglin II-305  
 Guo, Yongxin I-147
- Hai, Dan I-713  
 Han, Dapeng II-75  
 Han, Feng I-570  
 Hashimoto, Koichi II-378  
 Hashimoto, Tskuya II-505  
 Haynes, Paul S. II-717  
 Hei, Mo I-509  
 Herrera, Jose Luis Alarcon I-692  
 Horn, Joachim II-215  
 Hou, Jiaoyi II-296  
 Hou, Jung-Fu II-154  
 Hou, Lei II-378  
 Hou, Zengguang II-417  
 Hu, Huosheng II-659  
 Hu, Pinghua I-81  
 Hu, Xiangtao I-383  
 Huang, Dan I-171  
 Huang, Jian II-440  
 Huang, Xinhua I-558  
 Huang, Yong'an I-383  
 Hwang, Do-hyeong I-582
- Janaideh, Mohammad Al I-104  
 Jauw, Veronica Lestari II-612  
 Jeschke, Sabina II-44  
 Jia, Xinghua II-285  
 Jia, Zhenyuan II-285, II-565  
 Jiang, Dongsheng I-400  
 Jiang, Jiaxin II-390  
 Jiang, Li I-38  
 Jiang, Xianzhi II-463  
 Jin, Ming-He I-1, I-762  
 Ju, Zhaojie I-71
- Kagami, Shingo II-378  
 Kang, Zhan II-317  
 Kato, Naoki II-505  
 Kawano, Yuki II-357  
 Kawasue, Kikuhito II-357  
 Khoury, Mehdi I-680
- Ko, Kwang-Eun I-702  
 Kobayashi, Hiroshi II-505  
 Kouskouridas, Rigas II-705  
 Kubota, Naoyuki II-486, II-517, II-529  
 Kunze, Ralph II-44
- Lai, Leijie I-114  
 Lee, Hyun I-237  
 Lee, Jang Myung I-237  
 Lee, Jun seok I-237  
 Leong, Benny I-276  
 Li, Bolan I-372  
 Li, Fei I-477  
 Li, Gang II-305, II-317  
 Li, Jau-Rong I-266  
 Li, Jiting II-398  
 Li, Miao II-10, II-240, II-264  
 Li, Nan I-38, I-642, I-654  
 Li, Pengfeng II-417  
 Li, Pengzhi I-114  
 Li, Ping II-486  
 Li, Qingling II-417  
 Li, Shan II-22  
 Li, Shunchong I-81  
 Li, Xiuli I-300  
 Li, Yanan II-683  
 Li, Yangmin I-214, II-451  
 Li, Yanming I-642, I-654  
 Li, Yingsong II-227  
 Li, Youfu I-749  
 Li, Zheng I-246  
 Li, Zhijun I-171  
 Lii, Neal Y. I-1  
 Lin, Zhongqin II-252  
 Linsen II-754  
 Liu, Chang-Huan II-203  
 Liu, Chengliang I-195, I-642, I-654  
 Liu, Dehao II-305  
 Liu, Deshi II-305  
 Liu, Gequn I-204  
 Liu, Haining I-642, I-654  
 Liu, Hong I-1, I-38, I-570, I-762  
 Liu, Honghai I-71, I-680, II-486  
 Liu, Huaiyin I-329  
 Liu, Jiajun II-695  
 Liu, Jianqin I-257  
 Liu, Jing-Sin II-154  
 Liu, Lei I-204  
 Liu, Mei II-649  
 Liu, Ran I-310

- Liu, Shuangjun II-285  
 Liu, Wei II-285  
 Liu, Weiting I-477  
 Liu, Xinhui II-180  
 Liu, Xin-Jun I-318  
 Liu, Yi-Wei I-762  
 Liu, Yu I-570  
 Liu, Zhenglin I-183  
 Lu, Huimin II-105  
 Lumia, Ron I-630  
 Luo, Dezhi I-356  
 Luo, Jun II-649  
 Luo, Yaqin I-310  
 Luo, Yun II-742  
 Luo, Zi-rong I-499, I-509
- Ma, Hongxu I-736  
 Ma, Wentao I-519  
 Ma, Xiande II-56  
 Maestas, Dominic R. I-630  
 Mahmoud, Abeer II-166  
 Manabe, Takao II-97  
 Mao, Jianqin I-125, I-147  
 Martínez-García, Edgar I-453  
 Matson, Eric T. I-276  
 Mavrinac, Aaron I-692  
 Mayer, Marcel II-44  
 McKerrow, Phillip J. II-130  
 Meng, Tao I-724  
 Miao, Zhonghua I-195  
 Miki, Hiroyuki II-742  
 Min, Huasong I-183  
 Mu, Hua I-736
- Nagatomo, Satoshi II-357  
 Nguyen, Cory Q. I-276  
 Ni, Fenglei I-570  
 Nishiyama, Masahiro II-357  
 Niu, Peifeng II-368
- Okada, Tokuji II-166  
 Ou, Yongsheng I-548, II-191  
 Oyong, Arif Wicaksono II-612
- Pan, Minghua I-392  
 Parasuraman, S. II-612, II-624  
 Park, Gye-do I-237  
 Penders, Jacques II-717  
 Peng, Fangyu II-754  
 Peng, Fuyuan I-530
- Peng, Liling II-390  
 Peng, Yan II-649  
 Poignet, Philippe II-428  
 Ponnambalam, S.G. II-587,  
 II-599, II-636
- Qian, Jijie I-418  
 Qiao, Jinli I-257  
 Qiao, Jin-wei I-499
- Rakheja, Subhash I-104  
 Ren, Haoling II-730  
 Rönnau, Arne I-618  
 Ruan, Xiaogang II-87  
 Ruwadi, II-428
- Sarkodie-Gyan, Thompson I-288  
 Schamm, Thomas I-618  
 Schilberg, Daniel II-44  
 Shang, Jian-zhong I-499, I-509  
 Shang, Yonggang II-305  
 Shen, Xiong I-530  
 Shi, Hu I-329  
 Shi, Kun I-538  
 Shiota, Hiroyuki II-742  
 Sim, Kwee-Bo I-702  
 Smith, Anthony I-276  
 Spier, Eric I-288  
 Starr, Gregory I-630  
 Stefanini, Cesare I-477  
 Stuedemann, Stephan II-215  
 Su, Chun-Yi I-92  
 Sun, Haitao I-59  
 Sun, Jianzhong I-310  
 Sun, Jie I-59  
 Sun, Ronglei II-463  
 Sun, Wei I-310  
 Sun, Yuwen II-565  
 Sun, Zhenguo II-56, II-695
- Takagi, Toshiyuki II-742  
 Takeno, Takanori II-742  
 Tamura, Hiroki II-97  
 Tan, Dawei I-488  
 Tan, Guo yuan I-530  
 Tan, Min II-417  
 Tan, Qing I-348  
 Tang, Yandong I-600, I-609  
 Tanno, Koichi II-97  
 Tech, Ang Wei II-428

- Teh, Joshua II-636  
 Teng, Tao I-356  
 Tharin, Julien II-717  
 Thelen, Sebastian II-44  
 Tian, Bo I-442  
 Tian, Ke I-363  
 Tian, Yao-bin II-203  
 Toh, Chen Koon I-442  
 Torres-Córdoba, Rafael I-453  
  
 Veera Ragavan, S. II-636  
 Votelho, Wagner II-166  
  
 Wachs, Juan P. I-276  
 Wang, Aihui I-135  
 Wang, Bangguo II-285  
 Wang, Chaoqun I-409  
 Wang, Dongyun I-135  
 Wang, Hao II-252  
 Wang, Jingguo I-214, II-451  
 Wang, Jinsong I-318  
 Wang, Kai I-348  
 Wang, Licheng II-576  
 Wang, Lintao I-329  
 Wang, Qiyuan II-87  
 Wang, Shaojun II-576  
 Wang, Shuang II-398  
 Wang, Tianmiao I-183  
 Wang, Tongjian II-180  
 Wang, Xiangke II-75, II-105  
 Wang, Xi-ming II-203  
 Wang, Xinqing I-38  
 Wang, Xuezhzhu I-159  
 Wang, Xuyong I-195  
 Wang, Yongji II-440  
 Wang, Yu I-590  
 Wang, Yuhan II-553  
 Wang, ZhenYan I-125  
 Wang, Zhuo I-509  
 Wang, Zongyao I-724  
 Wei, Guowu I-430  
 Wei, Hongxing I-183  
 Wei, Lai II-659  
 Wood, John I-630  
 Wu, Chao I-318  
 Wu, Gongping II-227  
 Wu, Hongtao I-409  
 Wu, Jun II-440  
 Wu, Meiping I-736  
  
 Wu, Wenqi I-736  
 Wu, Xinyu I-548, II-191  
  
 Xia, Chunming I-590  
 Xia, Yi-min I-341, I-348, I-356  
 Xiang, Kangtai II-695  
 Xiao, hua II-227  
 Xiao, Zhihu I-530  
 Xie, Haibo II-730  
 Xie, Jinghua I-363  
 Xie, Ming I-442  
 Xie, Xiaoliang II-417  
 Xing, Kexin II-440  
 Xiong, Caihua II-463, II-541  
 Xiong, Youlun II-463  
 Xiu, CaiJing II-1  
 Xu, Feifei II-565  
 Xu, Guohua I-530  
 Xu, He I-488, II-142  
 Xu, Ke I-548  
 Xu, Wenfu I-570  
 Xu, Xiaolong I-530  
 Xu, Xiaoming I-204  
 Xu, Zhen II-329, II-339  
 Xu, Zijun I-348  
 Xue, Zhixing I-26  
 Xue, Yong I-499  
  
 Yaguchi, Aiko II-529  
 Yamashita, Yuki II-97  
 Yan, Rong II-754  
 Yan, Weixin I-400, I-519  
 Yan, Yonghua II-475  
 Yan, yu II-227  
 Yang, Dapeng I-38  
 Yang, Dawei I-363  
 Yang, Hejin I-442  
 Yang, Huayong II-730  
 Yang, Jianzhong II-754  
 Yang, Jun II-649  
 Yang, Lixin II-329  
 Yang, Ting II-22  
 Yang, Wenyu II-10, II-240, II-264  
 Yao, Yan-an II-203  
 Yao, Yu I-409  
 Yao, Zhenqiang II-276  
 Yap, Shee Cheng II-428  
 Yasuda, Gen'ichi II-32  
 Yi, Dong kyu I-237  
 Yin, Hang II-475

- Yin, Zhouping I-383  
 Yogeswaran, M. II-587  
 Yorita, Akihiro II-517  
 Yu, Huiying I-288  
 Yu, Naigong II-87  
 Yuan, Wenzhen II-671  
 Yun, Soh Chin II-624  
  
 Zeng, Xiangjin I-558  
 Zhang, Dan I-418  
 Zhang, Dingguo I-81  
 Zhang, Dinghua II-22  
 Zhang, Feng II-417  
 Zhang, Guoqing I-442  
 Zhang, Houxiang I-300  
 Zhang, Hui I-713, II-75, II-105  
 Zhang, Kai II-317  
 Zhang, Kui I-341  
 Zhang, Pu II-276  
 Zhang, Ruirui I-430  
 Zhang, Wenwen I-538  
 Zhang, Wenzeng I-13, I-47, I-59,  
     I-465, II-56, II-671  
 Zhang, XiaoJian II-541  
 Zhang, Xiaoping II-10, II-240, II-264  
 Zhang, Xu I-372  
 Zhang, Xuguang II-368  
 Zhang, Yanduo I-558  
 Zhang, Yang II-180  
 Zhang, Yitong I-257  
  
 Zhang, Yong-de II-405  
 Zhang, Yuan-Fei I-762  
 Zhang, Zhen I-226, II-390  
 Zhang, Zhenyu I-488, II-142  
 Zhao, Deyang I-13, I-465  
 Zhao, Jinfeng I-488  
 Zhao, Kai II-252  
 Zhao, Yan-jiang II-405  
 Zhao, Yanzheng I-400, I-519  
 Zhao, Yanzhi I-724  
 Zheng, Guilin I-159  
 Zheng, Lee Yi II-599  
 Zheng, Ruoyin II-398  
 Zheng, Zhiqiang I-713, II-105  
 Zhou, Chao I-538  
 Zhou, Fan II-345  
 Zhou, Hong I-590  
 Zhou, Hua II-296  
 Zhu, Guoli I-392  
 Zhu, Jianying I-409  
 Zhu, Jun I-590  
 Zhu, Limin I-114, I-372, II-553  
 Zhu, Linlin I-609  
 Zhu, Xiangyang I-81  
 Zhu, Xiaojin I-195  
 Zhu, Zongming I-356  
 Zhuan, Xiangtao I-159  
 Zöllner, J. Marius I-618  
 Zou, Jun II-296