Lecture Notes in Electrical Engineering 451

Kesheng Wang
Yi Wang
Jan Ola Strandhagen
Tao Yu *Editors*

Advanced Manufacturing and Automation VII

Springer

Lecture Notes in Electrical Engineering

Volume 451

"Lecture Notes in Electrical Engineering (LNEE)" is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer's high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer's other Lecture Notes series, LNEE will be distributed through Springer's print and electronic publishing channels.

More information about this series at http://www.springer.com/series/7818

Kesheng Wang · Yi Wang
Jan Ola Strandhagen · Tao Yu
Editors

Advanced Manufacturing and Automation VII

Springer

Editors
Kesheng Wang
Department of Mechanical
and Industrial Engineering
Norwegian University of Science
and Technology
Trondheim, Sør-Trøndelag Fylke
Norway

Jan Ola Strandhagen
Department of Mechanical
and Industrial Engineering
Norwegian University of Science
and Technology
Trondheim, Sør-Trøndelag Fylke
Norway

Yi Wang
School of Business
Plymouth University
Plymouth
UK

Tao Yu
Shanghai Second Polytechnic University
Shanghai
China

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-10-5767-0 ISBN 978-981-10-5768-7 (eBook)
https://doi.org/10.1007/978-981-10-5768-7

Library of Congress Control Number: 2017958627

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

IWAMA—International Workshop of Advanced Manufacturing and Automation—aims at providing a common platform for academics, researchers, practicing professionals, and experts from industries to interact, discuss trends and advances, and share ideas and perspectives in the areas of manufacturing and automation.

IWAMA began in Shanghai University 2010. In 2012 and 2013, it was held at the Norwegian University of Science and Technology, in 2014 at Shanghai University again, in 2015 at Shanghai Polytechnic University, and in 2016 at Manchester University. The sponsors organizing the IWAMA series have expanded to many universities throughout the world including Norwegian University of Science and Technology, Plymouth University, Manchester University, Shanghai University, Shanghai Polytechnic University, Xiamen University of Science and Technology, Tongji University, University of Malaga, University of Firenze, Stavanger University, The Arctic University of Norway, Shandong Agricultural University, China University of Mining and Technology, Indian National Institute of Technology, Donghua University, Shanghai Jiao Tong University, Changshu Institute of Technology, Dalian University, St. Petersburg Polytechnic University, Hong Kong Polytechnic University, Changzhou University, and China Instrument and Control Society. As IWAMA becomes an annual event, we are expecting that more sponsors from universities and industries will participate the international workshop as co-organizers.

Manufacturing and automation have assumed paramount importance and are vital factors for the economy of a nation and the quality of life. The field of manufacturing and automation is advancing at a rapid pace and new technologies are also emerging in the field. The challenges faced by today's engineers are forcing them to keep on top of the emerging trends through continuous research and development.

IWAMA 2017 takes place in Changshu Institute of Technology, China, September 11–12, 2017, organized by Changshu Institute of Technology, Plymouth University, and Norwegian University of Science and Technology. The program is designed to improve manufacturing and automation technologies for the next generation through discussion of the most recent advances and future perspectives,

and to engage the worldwide community in a collective effort to solve problems in manufacturing and automation.

Manufacturing research includes a focus on the transformation of present factories, toward reusable, flexible, modular, intelligent, digital, virtual, affordable, easy-to-adapt, easy-to-operate, easy-to-maintain, and highly reliable "smart factories". Therefore, IWAMA 2017 has mainly covered five topics in manufacturing engineering:

1. Industry 4.0
2. Manufacturing Systems
3. Manufacturing Technologies
4. Production Management
5. Design and optimization

All papers submitted to the workshop have been subjected to strict peer-review by at least two expert referees. Finally, 75 papers have been selected to be included in the proceedings after a revision process. We hope that the proceedings will not only give the readers a broad overview of the latest advances, and a summary of the event, but also provide researchers with a valuable reference in this field.

On behalf of the organization committee and the international scientific committee of IWAMA 2017, I would like to take this opportunity to express my appreciation for all the kind support, from the contributors of high-quality keynotes and papers, and all the participants. My thanks are extended to all the workshop organizers and paper reviewers, to CSLG, NTNU, and PU for the financial support, and to co-sponsors for their generous contribution. Thanks are also given to Lanzhong Guo, Ziqiang Zhou, Haishu Ma, Quan Yu, Zhe Li, Jin Yuan, and Jinghui Yang for their hard editorial work of the proceedings and arrangement of the workshop.

Plymouth, UK

Yi Wang
Chair of IWAMA 2017

Committees

Organized and Sponsored by
CSLG (Changshu Institute of Technology, China)
NTNU (Norwegian University of Science and Technology, Norway)
PU (Plymouth University, UK)

Co-organized by
SPU, Shanghai Polytechnic University
CU, Changzhou University
SHU, Shanghai University
TJU, Tongji University

Honorary Chairs
Prof. Minglun Fang, China
Prof. Kesheng Wang, Norway

General Chairs:
Dr. Yi Wang, UK
Prof. Jan Ola Strandhagen, Norway
Prof. Tao Yu, China

Local Organizing Committee
Prof. Lanzhong Guo (Chair)
Assoc. Prof. Ziqiang Zhou
Dr. Jian Wu

International Program Committee
Jan Ola Strandhagen, Norway
Kesheng Wang, Norway
Asbjørn Rolstadås, Norway
Per Schjølberg, Norway
Knut Sørby, Norway
Erlend Alfnes, Norway
Heidi Dreyer, Norway

Torgeir Welo, Norway
Leif Estensen, Norway
Hirpa L. Gelgele, Norway
Wei D. Solvang, Norway
Yi Wang, UK
Chris Parker, UK
Jorge M. Fajardo, Spain
Torsten Kjellberg, Sweden
Fumihiko Kimura, Japan
Gustav J. Olling, USA
Michael Wozny, USA
Wladimir Bodrow, Germany
Guy Doumeingts, France
Van Houten, Netherlands
Peter Bernus, Australia
Janis Grundspenkis, Latvia
George L. Kovacs, Hungary
Rinaldo Rinaldi, Italy
Gaetano Aiello, Italy
Romeo Bandinelli, Italy
Yafei He, China
Jawei Bai, China
Jinhui Yang, China
Dawei Tu, China
Minglun Fang, China
Binheng Lu, China
Xiaoqien Tang, China
Ming Chen, China
Xinguo Ming, China
Keith C. Chan, China
Xiaojing Wang, China
Jin Yuan, China
Yongyi He, China
Shili Tan, China
Ming Li, China
Chaodong Li, China
Cuilian Zhao, China
Chuanhong Zhou, China
Jianqing Cao, China
Yayu Huang, China
Shirong Ge, China
Jianjun Wu, China
Guijuan Lin, China
Shangming Luo, China
Dong Yang, China

Zumin Wang, China
Guohong Dai, China
Sarbjit Singh, India
Vishal S. Sharma, India

Secretariat:
Ziqiang Zhou
Jinghui Yang
Haishu Ma

Contents

How AI Affects the Future Predictive Maintenance: A Primer of Deep Learning 1
Kesheng Wang and Yi Wang

Application of Long Short-Term Memory Neural Network to Sales Forecasting in Retail—A Case Study 11
Quan Yu, Kesheng Wang, Jan Ola Strandhagen and Yi Wang

A Coarse-to-Fine Matching Method in the Line Laser Scanning System 19
Leilei Zhuang, Xu Zhang and Weiyan Zhou

Pose and Position Measurement in Dynamic Optical Coordinate Measure System 35
Yong Qi, Xu Zhang, Hao Wu and Wei Cheng

Design of Pin-Point Gate Injection Mould for Shells of Earplugs 53
Xian Feng, Min Yang and Min Zou

Research on the Principle of Automatic Correcting Machine 63
Haoyan Wang and Ziqiang Zhou

Development on Management System of Automated High-Rise Warehouse for Mid-Small Enterprises Based on Django 71
Chuanhong Zhou, Chao Dai, Pujia Shuai and Fei Qi

Underwater Image Enhancement by Fusion 81
Can Zhang, Xu Zhang and Dawei Tu

Restoration and Enhancement of Underwater Light Field Image 93
Wei Cui, Chen Li, Can Zhang and Xu Zhang

Optimizing Leather Cutting Process in Make-to-Order Production to Increase Hide Utilization 107
Giuseppe Fragapane, Oladipupo Olaitan, Erlend Alfnes and Jan Ola Strandhagen

A Study on Material Properties of Sealed Rubber Cylinder for Compressible Packer 115
Lixin Lu, Yue Qian, Guiqin Li and Peter Mitrouchev

Finite Element Analysis of Sealing Performance for Multi Cylinder Packer 123
Li-xin Lu, Tao Deng, Gui-qin Li and Peter Mitrouchev

Experimental Study on the Effect of Surface Texture on the Dynamic Performance of Journal Bearing 131
Jin Zhang, Guoping Li, Xiaojing Wang, Xiong Xin and Zikai Hua

The Effect of Temperature on Mechanical Properties of Polypropylene 143
Guiqin Li, Junjie Li, Jun Wang, Jiemin Feng, Qing Guo, Junlong Zhou and Peter Mitrouchev

The Effect of Extensometers on the Mechanical Properties of the Polypropylene Under Uniaxial Tensile Loading 151
Guiqin Li, Hao Chen, Jun Wang, Jiemin Feng, Qing Guo, Junlong Zhou and Peter Mitrouchev

Seal Property of Rubber Cylinder Shoulder in Packer 157
Li-xin Lu, Tao Deng, Gui-qin Li and Mitrouchev Peter

Data Acquisition and Storage Network Based on CAN-Bus 163
Lixin Lu, Jun Hua, Guiqin Li and Mitrouchev Peter

Effect of Pressure on Packer's Sealing Performance 169
Li-xin Lu, Lei Tang, Gui-qin Li and Peter Mitrouchev

Acquisition and Control Systems of Distributed Data Based on STM32 175
Lu Lixin, Shen Deshuai, Li Guiqin and Mitrouchev Peter

Introduction of Cyber-Physical System in Robotized Press-Brake Line for Metal Industry 181
Beibei Shu, Gabor Sziebig and Bjørn Solvang

Dynamic Balance System Design and Control for High-Branch Pruning Machine 187
Yang Li, Shuai Du, Xinxue Zhao, Ziru Niu and Jin Yuan

Braking Distance Monitoring System for Escalator 197
Xiaomei/XM Jiang, Shuguang/SG Niu and Lanzhong/LZ Guo

Design of Large Tonnage Elevator Sheave Block 207
Lanzhong Guo, Xiaomei Jiang and Yannian Rui

A Case Study of Tacit Knowledge Diffusion Object Preference in R&D Teams 219
Yang Shi

Strategic Framework for Manual Assembly System Design 229
Logan Reed Vallandingham, Jo Wessel Strandhagen, Erlend Alfnes and Jan Ola Strandhagen

The Research and Development of Preventing the Accidental Movement of the Elevator Car Safety Protection Device 239
ShuGuang Niu, Huiling Liu, Guokang Chen and Haizhou Zhang

Ultrasonic Sensing System Design and Accurate Target Identification for Targeted Spraying 245
Xinxue Zhao, Yang Li, Xuemei Liu, Ziru Niu and Jin Yuan

Fault Location in Power System Based on Different Modes of Traveling Wave and Artificial Neural Network 255
Chenglei Liu, Ke Bi and Rui Liang

Review of Technology for Strengthening Effect of Fe Element in Al Alloys 265
De-Qin Sun, Zi-Qiang Zhou, Su Zhu and Pei-Jun Wu

Characterization the Fatigue Life of the Flame Thermal Spray Parts Applying the Power Function Energy Method 275
Zhi-Ping Zhao, Chao Wang and Xin-Yong Li

Design of Poultry Egg Quality Detection System Based on LABVIEW and PLC 285
Yan Liu and Rui-Jie Kong

Thermal Deformation of a Vertical Plate According to Various Plastic Deformation Regions 295
Teng Yao, Lanzhong Guo and Chao Wang

The Comparison of Different Models of Government Purchasing Home-Based Aged-Care Services 305
Guoping Zhang

Study on Value Increment of Channel Value Chain Between Manufacturer and Distributor-Based on Game Theory 313
Junwei Ma and Jianhua Wang

The Study on Evaluation Index System of Restructuring Construction Industry Under the Green Development Model 325
Fang Liu and Jie Zhao

Correcting Strategy for Distortion of Ring Part Based on Genetic Algorithm 335
Ziqiang Zhou, Ke Be and Haoyan Wang

Human Centered Automation System for ELV Disassembly Line 347
Guohong Dai and Ziqiang Zhou

Rolling Bearing Fault Diagnosis Using Deep Learning Network 357
Shenghao Tang, Yuqiu Yuan, Li Lu, Shuang Li, Changqing Shen and Zhongkui Zhu

An Automatic Feature Learning and Fault Diagnosis Method Based on Stacked Sparse Autoencoder 367
Yumei Qi, Changqing Shen, Jie Liu, Xuwei Li, Dading Li and Zhongkui Zhu

Application of Intelligent Hanging Production System in Garment Industry 377
Jishu Zhang, Yingzi Zhang and Peiguo Wang

Objective Evaluation of Clothing Modeling Based on Eye Movement Experiment 383
Yingzi Zhang and Jishu Zhang

Simulation Analysis of Rotor Indirect Field Oriented Vector Control System for AC Induction Motor in Low Speed Electric Vehicles 391
Kang Wang, Qingzhang Chen and Zhengyi Wang

Vision Based Quality Control and Maintenance in High Volume Production by Use of Zero Defect Strategies 405
Håkon Raabe, Odd Myklebust and Ragnhild Eleftheriadis

A RFID Based Solution for Managing the Order-Picking Operation in Warehouse 413
Haishu Ma, Jinghui Yang and Kesheng Wang

Improving the Decision-Making of Reverse Logistics Network Design Part II: An Improved Scenario-Based Solution Method and Numerical Experimentation 421
Hao Yu and Wei Deng Solvang

Improving the Decision-Making of Reverse Logistics Network Design Part I: A MILP Model Under Stochastic Environment 431
Hao Yu and Wei Deng Solvang

Predictive Maintenance for Synchronizing Maintenance Planning with Production 439
Harald Rødseth, Per Schjølberg, Markus Wabner and Uwe Frieß

Recognition of Garlic Cloves Orientation Based on Machine Vision 447
Shuai Du, Yang Li, Xuemei Liu and Jin Yuan

Metal Penetration in Additively Manufactured Venting Slots for Low-Pressure Die Casting 457
Even Wilberg Hovig, Knut Sørby and Per Erik Drønen

Analysis and Impact Assessment in Sustainable Industrial and Infrastructural Development Projects 469
Lea Kristine Myklebust and Odd Myklebust

Lightweight Design of LED Street Lamp Based on Response Surface Method 479
Mengjia Qi, Lilan Liu, Renfei Ma and Yi Wang

Industry 4.0 and Cyber Physical Systems in a Norwegian Industrial Context 491
Ragnhild J. Eleftheriadis and Odd Myklebust

Mechanical Analysis of a Customized Hand Orthosis Based on 3D Printing 501
Jiawei Wu, Cuilian Zhao, Yexiao Liu and Shuangchi Ma

Fault Classification and Degradation Assessment Based on Wavelet Packet Decomposition for Rotary Machinery 509
Zhe Li, Viggo Gabriel Borg Pedersen, Kesheng Wang and Yafei He

Research on Assembly and Disassembly of Reducer with the Combination of Virtual and the Actual Reality 517
Wenhua Zhu, Bao Cai, Kunju Shi and Sai Liu

Equipment Condition Monitoring System Based on LabVIEW 531
Zhongwei Rao, Bowen Feng, Lilan Liu and Yi Wang

The Algorithm Knowledge Base for Steel Production Process Optimization 541
Shuai Li, Lilan Liu, Yiping Wu and Yi Wang

Study on the Friction Characteristics of Two Polytetrafluoroethylene (PTFE) Coatings 553
Jian Wu, Te Li and Lanzhong Guo

A Hybrid Nested Partitions Method for Bi-objective Job Shop Scheduling Problem Considering Energy Consumption and Makespan 563
Mei Dai, Xin Li and Limin Mao

Research on the Framework of Quality Prediction in Intelligent Manufacturing 573
Min Ji

Deep Learning Approach to Multiple Features Sequence Analysis in Predictive Maintenance 581
Jin Yuan, Kesheng Wang and Yi Wang

Sorting System of Robot Based on Vision Detection 591
Qin Qin, Dongdong Zhu, Zimei Tu and Jianxiong Hong

The Challenges and Promises of Big Data—An Engineering Perspective 599
Yi Wang

A Modified Teaching and Learning Based Optimization Algorithm and Application in Deep Neural Networks Optimization for Electro-Discharge Machining 605
Chen Wang, Baorui Li, Yi Wang, Kesheng Wang and Shenghuai Wang

A Study on a Novel Application of Eye Tracking Technology in Product Customization 617
Baorui Li, Yi Wang, Kesheng Wang, Jinghui Yang and Lilan Liu

Big Data Analysis in Click Prediction 629
Lapo Chirici and Kesheng Wang

About the Editors

Prof. Kesheng Wang holds Ph.D. in production engineering from the Norwegian University of Science and Technology (NTNU), Norway. Since 1993, he has been appointed as Professor at the Department of Mechanical and Industrial Engineering, NTNU. He is Director of the Knowledge Discovery Laboratory (KDL) at NTNU at present. He is also an active researcher and serves as Technical Adviser in SINTEF. He was elected as a member of the Norwegian Academy of Technological Sciences in 2006. He has published 19 books, 10 book chapters, and over 240 technical peer-reviewed papers in international journals and conferences. Professor Wang's current areas of interest are intelligent manufacturing systems, applied computational intelligence, data mining and knowledge discovery, swarm intelligence, condition-based monitoring, and structured light systems for 3D measurements and RFID, Predictive maintenance and Industry 4.0.

Dr. Yi Wang obtained his Ph.D. from Manufacturing Engineering Center, Cardiff University in 2008. He is Lecturer in Business School, Plymouth University, UK. Previously, he worked in the Department of Computer Science, Southampton University and at the Business School, Nottingham Trent University and at the School of Material, Manchester University. He holds various visiting lectureships in several universities worldwide. He has special research interest in supply chain management, logistics, operation management, culture management, information systems, game theory, data analysis, semantics and ontology analysis, Neuromarketing and design, and decision analysis. He has published more than 30 technical peer-reviewed papers in international journals and conferences. He co-authored two books: Operations Management for Business and Data Mining for Zero-defect Manufacturing.

Prof. Jan Ola Strandhagen is Professor at Department of Mechanical and Industrial Engineering, the Norwegian University of Science and Technology (NTNU). He holds Ph.D. in Production Engineering from NTNU (1994). He was the Research Director of the research center SFI NORMAN at SINTEF. His research has focused on production management and control, logistics, manufacturing economics, and strategies. He has managed and performed R&D projects in

close collaboration with a wide variety of Norwegian companies and participated as researcher and project manager in several European projects.

Prof. Tao Yu is President of Shanghai Second Polytechnic University (SPU), China and Professor of Shanghai University (SHU). He received his Ph.D. from SHU in 1997. He is a member of the Group of Shanghai manufacturing information and a committee member of the International Federation for Information Processing IFIP/TC5. He is also Executive Vice President of Shanghai Science Volunteer Association, and Executive Director of Shanghai Science and Art Institute of Execution. He managed and performed about 20 national, Shanghai, enterprises commissioned projects. He has published hundreds of academic papers, of which about thirty were indexed by SCI, EI. His research interests are mechatronics, computer integrated manufacturing system (CIMS), and grid manufacturing.

How AI Affects the Future Predictive Maintenance: A Primer of Deep Learning

Kesheng Wang and Yi Wang

Abstract We are in the midst of the industry 4.0 or the fourth industry revolution, a transformation revolving around intelligent sensors, machines, networks and business. Some of the newer concepts are overwhelming by their impact, and transformational technologies are just the tip of the iceberg. Artificial Intelligence, mainly Computational Intelligence will greatly affect future human's life, economics, business, industries and even political systems. In this paper, we only discuss about the impact of AI to future predictive maintenance, which is an important parts of future advanced production systems. Specially we focus on Deep Learning (DL) technology, which is one branch of Artificial Neural Networks (ANN) and try to answer some questions on what DL is and why we are interested in applying DL in predictive maintenance strategy today.

Keywords Industry 4.0 · Predictive maintenance · AI · Computational intelligence Artificial neural network · Deep learning

1 Introduction

It has been pasted almost 60 years since Artificial Intelligence (AI) was defined as an academic discipline of computing science. After 60 years development, AI is still a hot research field, but the contents of AI has been changed a lot. Leaders of the world's most influential technology firms, Google, Microsoft and Facebook, are demonstring their interests in AI. AI together with other technologies, such as sensors, IOT, cloud computing, big data, data analysis, and smart decision-making

K. Wang
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology (NTNU), NO 7491 Trondheim, Norway
e-mail: Kesheng.wang@ntnu.no

Y. Wang (✉)
School of Business, Plymouth University, Plymouth, UK
e-mail: Yi.wang@plymouth.ac.edu

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_1

are driving a new industry revolution, which we called as Industrial 4.0 or the fourth industry revolution. The use of smart and advanced technologies could automate near half of all existing work activities in the world and it save some $16 trillion in wages by replacing human operators. It was predicted that revenues from AI by 2020 will be increased more than $47 billion. Industry will invest in many promised fields, such as smart factory, smart logistics, medical diagnosis system, big data and decision management and fraud analysis in business [1].

AI has plays an important role in industry 4.0. Predictive maintenance is a branch of Industry 4.0, which can be defined as a set of activities that monitor changes of condition of processes or equipment in order to make correct and quick maintenance decision for maximizing the service life of equipment. Our intelligent predictive maintenance solution (Fig. 1) opens up innovative new possibilities for companies. The functions of the predictive maintenance system can be described as the following: a set of data generated by Cyber-physical systems (CPS) and transmitted by Internet of Things (IOT) is used to monitor machine/process condition and then it is automatically analysed in order to pick up any rules or patterns that indicate a possible fault through Data Mining systems for making a decision. By Internet of service (IOS), this decision can be used to allow giving alarm early and correction parameters. This innovative solution is called Industry 4.0 for Intelligent Predictive Maintenance (*IPdM*).

Our goal for this paper is to answer some questions on AI. For example, which is AI technology or what type of algorithm important for predictive maintenance? why it is so poplar now in engineering? How AI affects the future predictive maintenance strategy.

We organize this paper in 6 sections. Section 1 introduces the AI technology and intelligent predictive maintenance (IPdM) system. The fascinating Deep Learning

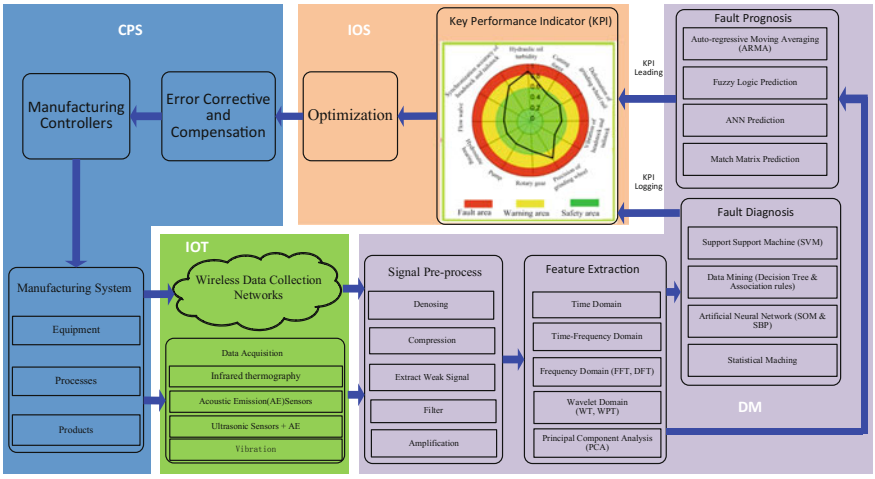


Fig. 1 The framework of intelligent predictive maintenance (*IPdM*) systems in industry 4.0 [8]

conception and the relationship with AI, CI, and Artificial Neural Networks are described in Sect. 2. How Deep Learning affects predictive maintenance strategy is presented in Sect. 3. In Sect. 4, we will discuss why AI is attracting so many academic researchers and practical engineers in the world to work with now. The conclusions and future development is summarised in Sect. 5.

2 Deep Learning

2.1 Artificial Intelligence

Artificial Intelligence has been part of our imaginations and issues in research labs since computers became available and popular in real life. To be more specific, its in 1956, when a handful of computer scientists rallied around the term at the Dartmouth Conferences and birthed the field of AI. The idea is to make machines be able to carry out tasks in a way that we would consider "smart" or "intelligent". The concept is very broad from today's perspective. Based our understanding, the main classification of AI can be shown in Fig. 2. AI can be divided two types: Computational Intelligence (Numerical AI or soft computing). Computational Intelligence can be further classified several groups: 1. Artificial Neural Networks (ANN); 2. Fuzzy Logic Systems (FLS); 3. Evolutional Computing (EC); and 4. Swarm Intelligence (SI).

2.2 Computational Intelligence

Computational Intelligence (CI) [2] or Soft Computing (SC) [3] can be called Numerical AI (verse Symbolic AI) has become a rapid developing field in computer science and been an advanced information processing technology. Wang [4]

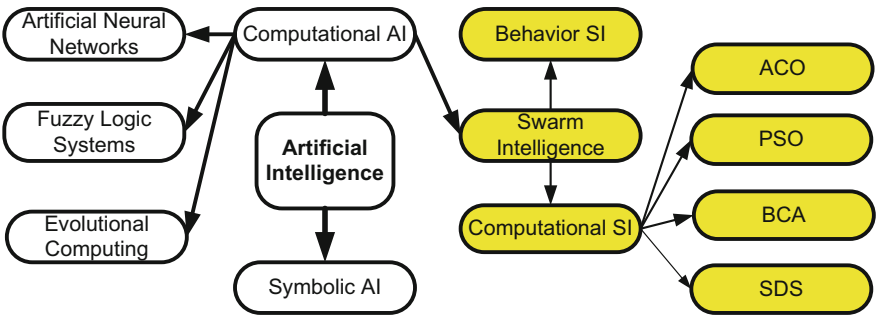


Fig. 2 The classification of AI [9]

classified CI into several groups, such as Artificial Neural Networks (ANN), Fuzzy Logic Systems (FLS) and Evolutionary Computing (Genetic Algorithms). The human brain is able to process data and information quickly and accurately because of its biological network structure and approximate reasoning ability. Both the functions of Artificial Neural Networks and Fuzzy Logic Systems are based on the mechanisms of human biological brain (Fig. 3). The ANN simulates physiological features of the human brain, and has been applied for non-linear mapping by a numerical approach. The FLS simulates psychological features of the human brain, and has been applied for linguistic translating by the use of membership functions.

2.3 *Artificial Neural Networks (ANN)*

The basic principle of ANN has been describe in Sect. 2.2. Because the functions of Artificial Neural Networks are based on biological neural network structure of humans brain, they are able to analyze complex/complicated problems and to make correct decisions. The powerful and paralleled computing architecture of ANN demonstrate their learning abilities. ANN can learn from history data, previous examples, and experience to do modeling, classification, clustering and predication. If the environment changes, ANN are able to modify their behavior. There are many ANNs application areas such as dynamic system modeling, pattern recognition, data classification, business predication, and multi-dimension mapping and novelty detection.

ANN uses artificial, software-based computing technologies that approximate the function of neurons in a brain are connected together. They form a 'neural network' which receives an input; analyses it; makes a determination about it and is

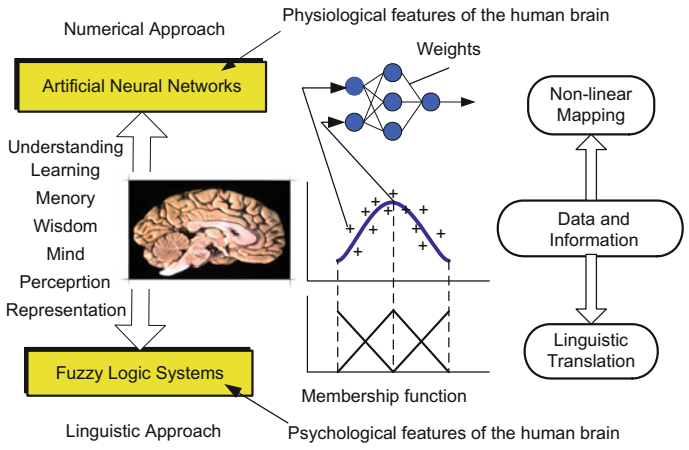


Fig. 3 ANNs and FLS based on the mechanism of the human brain [4]

informed if its determination is correct. If the output is wrong, the connections between the neurons are adjusted by the algorithm, which will change future predictions. Initially the network will be wrong many times. But as we feed in millions of examples, the connections between neurons will be tuned so the neural network makes correct determinations on almost all occasions. Practice makes (nearly) perfect.

2.4 Deep Learning (DL)

From the view of structure, DL is on kind of ANNs or we can say it is the application of Artificial Neural Networks (ANNs) to learning tasks that contain more than one hidden layer. The types of DL approaches have been growing increasingly richer, encompassing variety of Artificial Neural Networks with multiple processing layers, unsupervised or supervised learning algorithms. There exist about more than 10–15 types of DL algorithms. We present 4 types of DL, which are mostly used in predictive maintenance field: 1. Deep Feedforward Networks (DFN); 2. Long Short-Term Memory (LSTM); 3. Convolutional Networks (CN); 4. Deep Belief Networks (Fig. 4). We will not discuss the details of the structures and training process, which can be find in the MIT book [5].

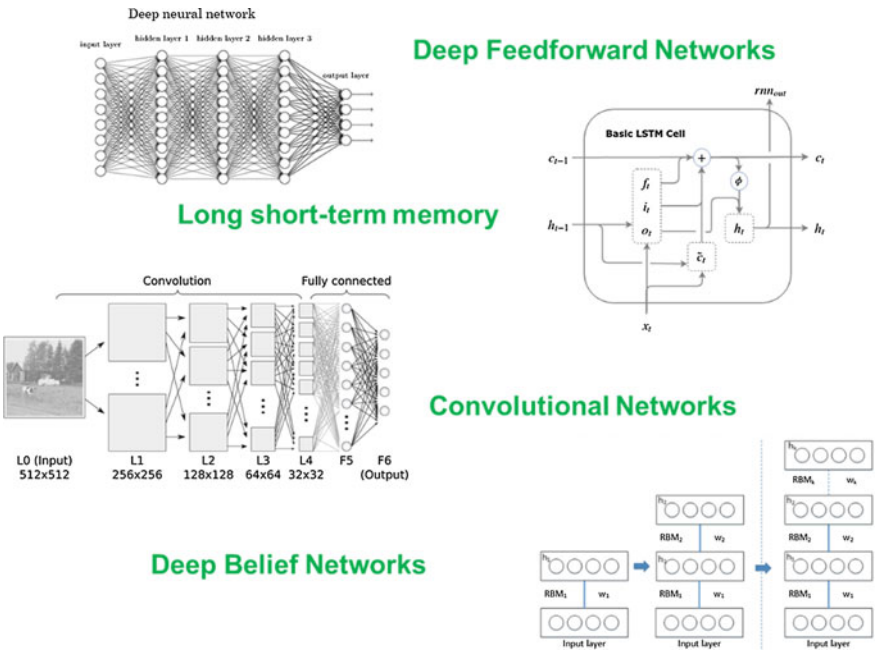


Fig. 4 The 4 types of deep learning networks

Deep learning is useful because it undertakes both the tasks of feature specification and optimization. Deep learning uses the same principles of ANNs, which approximate the function of neurons in a brain, are connected together. Using this approach, we can do the following tasks:

- Speech and image recognition;
- Nature language translation in real-time;
- Voice control;
- Prediction of genetic variation;
- Customer relationship management;
- Medical diagnosis;
- Prediction of machine health condition; and more.

The main difference between ordinary ANNs and Deep Learning Algorithms is that in DL the features are not designed by developers but learned/extracted from data itself through a generalized self-learning procedure. Deep learning is able to offer us a more suitable and convenient way for treating feature extraction problem.

However, DL is not well suited to every problem. It typically requires large data sets for training. It has a 'black box' problem—it can be difficult to know how a neural network developed its predictions. We need to improve our techniques to deal with industry big data.

There are many commercial applications driven by DL technology. Best example is that Google has developed their AlphaGo system based on the Deep Learning Algorithm by learning the game of Go well enough to beat a professional Go player [6]. It make people to focus more on AI and DL technologies.

3 Deep Learning for Predictive Maintenance

Deep learning has proven to show superior performance in certain domains such as object recognition and image classification. It has also gained popularity in domains such as finance where time-series data plays an important role. Predictive Maintenance is also a domain where data is collected over time to monitor the state of an asset with the goal of finding patterns to predict failures which can also benefit from certain deep learning algorithms.

More recently, DL, as a latest research area of ANN, has accelerated its application in maintenance and service. In the research area of predictive maintenance, conventional ANN methods have proved their values in fault diagnosis, e.g., BP supervised ANN and SOM unsupervised ANN perform well in the fault detection and classification, however, when the target condition is beyond history data, they don't have the ability to predict potential failures. This problem is common and challenging since in many situations, one may not have the data under faults, some machines may run several years without any failures). However, it may exist potential faults that may occur one day, and when it happens, it may cause terrible disasters in both economy and personal safety.

An important part of modeling a failure predictor is selecting or constructing the right features, i.e. selecting existing features from the data set, or constructing derivative features, which are most suitable for solving the learning task.

Traditionally, the features are selected manually, relying on the experience of process engineers who understand the physical and mechanical processes in the analyzed system. Unfortunately, manual feature selection suffers from different kinds of bias and is very labor intensive. Moreover, the selected features are specific to a particular learning task, and cannot be easily reused in a different task (e.g. the features which are effective for predicting failures in one production line will not necessarily be effective in a different line).

We have developed a general framework for predictive maintenance based on DL. We expect that it may overcome the disadvantage of ANN in predictive maintenance mentioned above. Figure 5 shows the framework, which has been use for a case: backlash prediction in a machine center.

The framework have been tested in several projects: Green Monitoring, Monitor X and rotational machinery rig in KDL lab at NTNU. Deep Learning techniques investigated in the Green Monitor project offer an alter-native to manual feature selection. It refers to automatically extract features from the raw data that are most suitable for solving a particular learning task. Predictive Maintenance can benefit from such automatic feature extraction to reduce effort, cost and delay that are associated with extracting good features.

One interesting commercial application of DL for predictive maintenance is that 3DSignals (ultrasonic sensors), relies on the DL algorithm to monitor sawing machines, learn the audio patterns from troubled machines. Then the system can

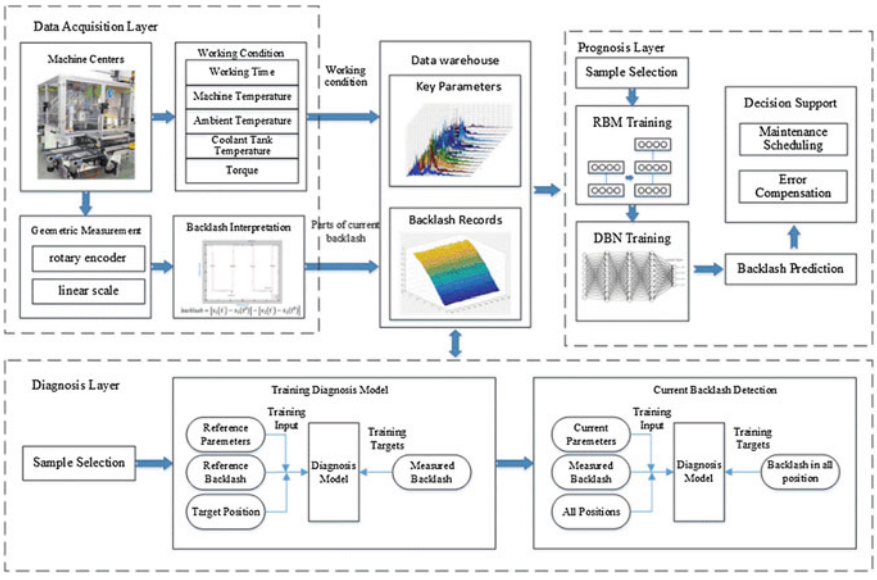


Fig. 5 A framework of predictive maintenance based on deep learning

predict failures of machines to realize predictive maintenance strategy [7]. 3DSignals has already begun talking with us to use the deep learning service to detect cavitation problems in hydropower plant.

4 Discussions

AI research began in the 1950s, after repeated false dawns, why is now the inflection point? The effectiveness of AI/CI/ANN/DL has been improved due to the innovation of several new algorithms, greater availability of data collection, better hardware to train them and cloud-based services to catalyse their adoption among developers. The main reasons could be listed as the following:

- Algorithms have been improved.

While deep learning is not new—the specification for the first effective, multi-layer neural network was published in 1965—evolutions in deep learning algorithms during the last decade have transformed results.

- Extensive data can be collected easily.

The neural networks used for deep learning typically require large data sets for training—from a few thousand examples to many millions. Fortunately, data creation and availability has grown exponentially. Beyond increases in the availability of general data, specialist data resources have catalysed progress in machine learning. ImageNet, for example, is a freely available database of over 10 million hand-labelled images. Its presence has supported the rapid development of object classification deep learning algorithms.

- Cloud services has been developed.

Developers' use of deep learning is being catalysed by the provision of cloud-based deep learning infrastructure and services from the industry's leading cloud providers. Many companies offer cloud-based infrastructure to reduce the cost and difficulty of developing Deep learning capabilities.

- Specialised hardware is possible.

Graphical Processing Units (GPUs) are specialised electronic circuits that are slashing the time required to train the neural networks used for deep learning. When combined with software development kits tuned for widely used deep learning frameworks, the improvements in training speed can be even greater.

- Interest and entrepreneurship has increased.

The public's interest in AI has increased in the last five years, with a still greater increase in the number of investments in AI companies by venture capital firms. We have entered a virtuous circle, in which progress in Deep learning is attracting investment, entrepreneurship and awareness.

5 Conclusions

The benefits of AI including CI, ANN and DL will be numerous and significant in engineering and industries. The development and research in AI can be useful for develop industry 4.0 scenario. In this paper, we focus on the description of a prime of Deep Learning and predictive maintenance. The integration of DL into predictive maintenance system is the key successful factor for industries. The biggest initial win with predictive maintenance is cost savings. But after implementing a larger, more robust, and more mature predictive maintenance strategy, larger opportunities begin to open from a business perspective, and high-value assets can bring in some additional revenue instead of just being costs. Predictive maintenance also lends itself to the future of AI, where operations will be entirely self-maintenance with very little human interaction whatsoever. AI in the predictive maintenance space would go one step beyond the steps discussed above, which would still require some manual analysis of models and outputs.

These systems will watch thousands of variables and apply deep learning to find information that could otherwise be undetected that might lead to failure. Ultimately, predictive maintenance isn't so far off from AI, and businesses that get started with predictive maintenance programs now will be well-poised as market leaders in the future.

References

1. https://www.brookings.edu/blog/brown-center-chalkboard/2017/04/11/surfing-the-4th-industrial-revolution-artificial-intelligence-and-the-liberal-arts/?
2. Bezdek JC (1992) On the relationship between neural networks, pattern recognition and intelligence. Int J Approx Reason 6:85–107
3. Jang JSR, Sun CT, Mizutani E (1997) Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence. Prentice Hall, Upper Saddle River, NJ, 07458. ISBN 0-13-261066-3
4. Wang K (2007) Applied computational intelligence in intelligent manufacturing. International series on natural and artificial intelligence, vol 2, 2nd edn. Advanced Knowledge International Ltd., Australia, p 454. ISBN 978-0-9751004-9-3
5. http://www.deeplearningbook.org/
6. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V (2016) Mastering the game of Go with deep neural networks and tree search. Nature 529(7587):484–489. https://doi.org/10.1038/nature16961. ISSN 0028-0836, PMID 26819042
7. http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/deep-learning-ai-listens-to-machines-for-signs-of-trouble
8. Wang K (2016) Intelligent predictive maintenace (IPdM) system - industry 4.0 scenario. In: Wang K, Wang Y, Strandhagen JO, Yu T (eds) Advanced Manufacturing and Automation V. WIT Transaction on Engineering Science, vol 113. pp 259–268, ISBN 978-1-78466-169-4
9. Wang Y, Liu L, Wang K (2012) Swarm intelligence (SI) for decision support of operations management—methods and applications. In: Dargam F, Delibasic B, Hernández JE, Liu S, Ribeiro R, Zaraté P (eds) Proceedings of the EWG-DSS livepool 2012 workshop, Liverpool, UK

Application of Long Short-Term Memory Neural Network to Sales Forecasting in Retail—A Case Study

Quan Yu, Kesheng Wang, Jan Ola Strandhagen and Yi Wang

Abstract Sales forecasting is an important task for managers to make replenishment according to historical sales. A flexible and easy to use forecasting solution will benefit retailers from loss of sale, over supply and merchandise waste. Deep learning is a popular topic in many fields in recent years. This paper tests a long short-term memory (LSTM) recurrent neural networks (RNN) on 45 weeks point of sale (POS) data of 66 products without considering the impact of seasonality and promotions. One fourth of products have a relatively low forecasting error, which validates the feasibility of the LSTM network to some degree.

Keywords Deep learning · LSTM · RNN · Sales forecasting

1 Introduction

Sales forecasting is a process to estimate future sales according to historical sales and other relevant information, which enables a company to predict short-term and long term business and make corresponding decisions. In retail industry, sales forecasting is important for inventory control, supply chain management, replenishment, etc. A good forecast that fits the consumption demands can help retailers to extend profits, promote products with respect to consumption patterns, and control the safety stock without being excessive. However, sales forecast methodologies are typically more scientific than an intuitive chart, although different forecasting models are available in various software, forecast executives in companies would not risk tuning the parameters for a potential better forecast [1]. On the other side, machine learning develops dramatically in recent years and contributes in both

Q. Yu (✉) · K. Wang · J. O. Strandhagen
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, Norway
e-mail: Quan.yu@ntnu.no

Y. Wang
School of Business, Plymouth University, Plymouth, UK

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_2

industry fields and people's daily life, with relatively simpler parameter settings as an alternative to traditional model-based methods.

A big branch of machine learning is neural networks having decades of development and many variations. The Long Short-Term memory (LSTM) neural network is a sort of recurrent neural network (RNN) proposed in 1997 to address the problem of insufficient, decaying error backflow in RNN training [2]. However, research interests on LSTM have been slowly increasing until 2014 as shown in Fig. 1. Publications are dramatically increased along with the application of LSTM to natural language processing by well-known companies such as Google [3]. Since then, LSTM has been applied in fields including robot control [4], speech recognition [5], time series forecasting [6], etc.

Because of the special structure, an LSTM is able to learn from historical information to classify, process and predict time series and reflects important events. This paper tests an LSTM network on real sales data of 66 products collected from a retailer during 45 weeks. As a general experiment, seasonality and promotions in practical are not under consideration. The results reflect the potential of implementing LSTM network in retail sales forecasting.

2 Long Short-Term Memory Neural Network

Similar to neurons in standard neural networks, the central part in a LSTM architecture is a memory cell which can maintain its state over time, and non-linear gating units which regulate the information flow into and out of the cell [7]. Comparing with traditional RNN, an LSTM neural network uses memory units with gates rather than neurons to establish connections between inputs and outputs [8]. Figure 2 shows a standard LSTM memory cell, where

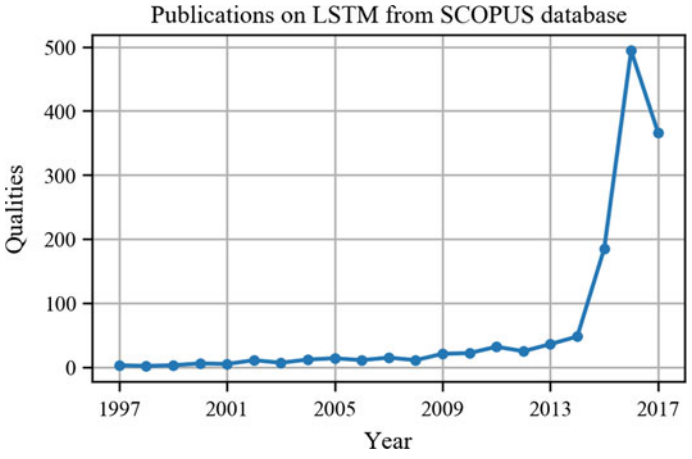


Fig. 1 Trend of research on LSTM

x_t is the input x and time t,
h_t is the output with respect to x_t,
h_{t-1} is the output of previous memory cell, also called hidden state,
C_t is the cell state while C_{t-1} is the cell state of the previous cell.

Within the memory cell, the input x_t will be processed together with the previous output h_{t-1} by three steps respectively, i.e. forget step, input step and output step.

- The forget step contains a sigmoid function, which returns a value between 0 and 1, while 0 means no information pass and 1 means all information pass.
- The input step has a sigmoid function and a tanh function, with the purpose of adding new information according to x_t and h_{t-1}.
- The output step is another sigmoid function to decide how much information is included in the output h_t.

Additionally, the cell state C_t is updated according to the previous cell state C_{t-1}, the output of the forget step for h_{t-1} and x_t, and the output of input step for h_{t-1} and x_t, thereafter feed to the output h_t and the next cell as a reference.

By stacking memory cells, information of original input $\boldsymbol{x}$ is kept in the final output $\boldsymbol{y}$ to some degree, carried by cell state $\boldsymbol{C}$ and cell by cell output $\boldsymbol{h}$. To be noticed, the stack of the memory cells can mimic a time series. Based on Fig. 2, correspondence between a time series and an LSTM network is shown in Fig. 3. Suppose we use an LSTM network with three memory cells for time series forecasting and $\boldsymbol{x}$ is a time series, cells with the same internal structure are stacked and each of them holds an input x_i sequentially. When an x_i is input to a memory cell, it is processed and output to be the cell state C_i and hidden state h_i, as inputs for the next cell together with next x_{i+1}. The last hidden state h_3 is taken as the final output y_4 corresponding to x_4. By sliding this three cells LSTM network along the time

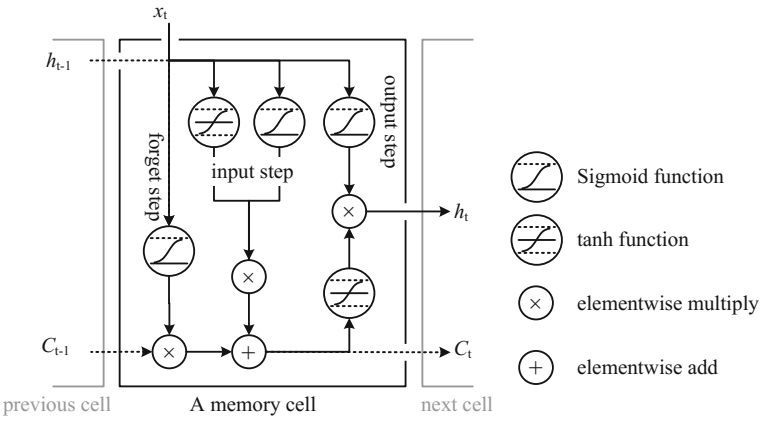


Fig. 2 An LSTM memory cell

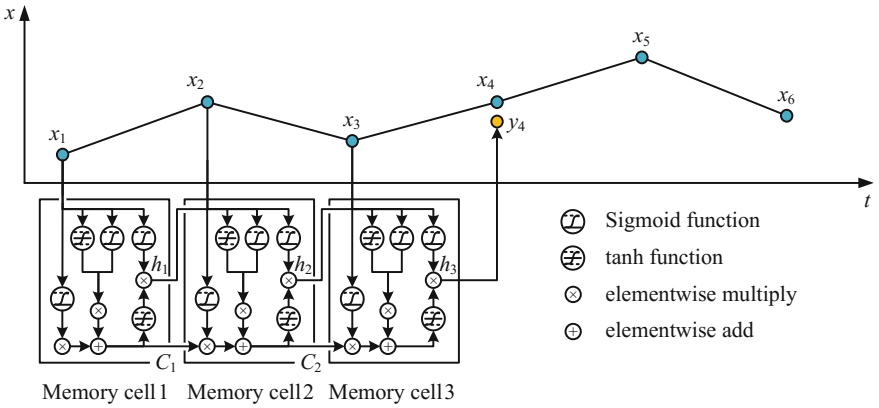


Fig. 3 A three cells LSTM network for time series forecasting

axis and aligning to the time stamp, every three inputs x will have the fourth x as the output, which can be used as training data to train the network and fix the parameters, thereafter to be ready for the forecasting later on.

3 Experiment and Results

In this paper, the LSTM network is established in python environment with Keras deep learning library running on top of TensorFlow library developed by Google. The sales data is of 66 products in 45 weeks in 2015, covering milk products, fruits and vegetables, meat products, etc. The forecasting is performed in week level, while sales of four consecutive weeks are used to forecast the sale of the fifth week. Accordingly, the LSTM network is set to have four memory cells. Raw data of each product is sequentially divided in the ratio 2:1 to be training data and test data. In other words, the data of first 30 weeks is taken as training data, while the latter 15 weeks data is used as test data. Furthermore, for better evaluating the performance of LSTM network, the sales values are scaled down between [0, 1] in the test so that they are normalized and 10 experiments are performed for each product and the average mean square error (MSE) is calculated as the final performance measure. For each product, the work flow is shown in Fig. 4. An example of the experiment results is shown in Fig. 5. There is no significant pattern in raw data and no large bias between raw data and forecasted values at the beginning, the forecasting accuracy is decreased along with a fluctuation of historical data at time

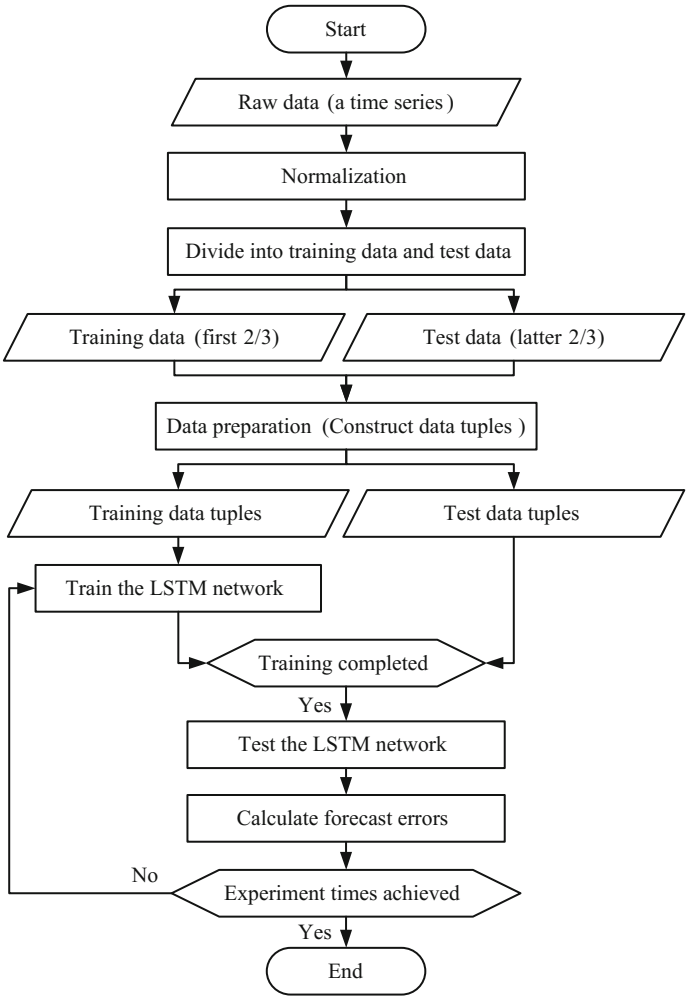


Fig. 4 Work flow of the experiment of one product

stamp 6. Regarding all products, 10 forecasting experiments are performed for each product and the final results are shown in Fig. 6, measured by average MSE values. There are about one fourth products having an MSE lower than 0.3, which means the forecasting accuracy for other products is not good enough.

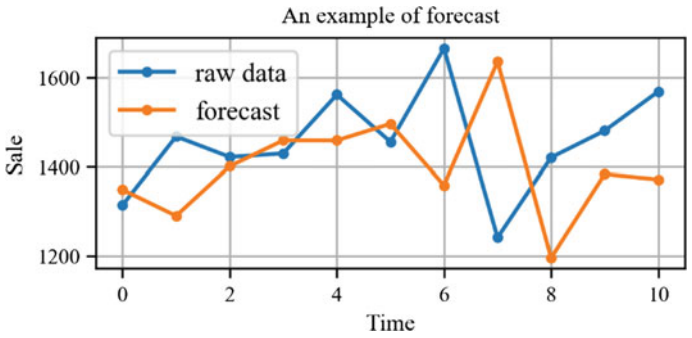


Fig. 5 A forecasting example from the results

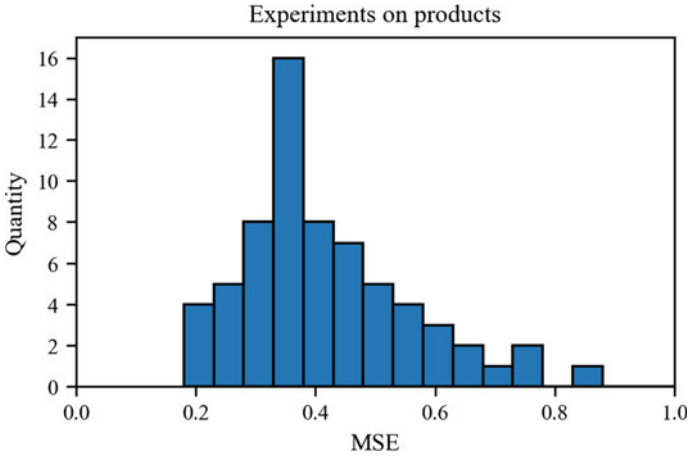


Fig. 6 Experiment results on products

4 Discussion and Conclusions

Sales data in retail industry can be seen as a sort of time series, which have been researched for decades. Various models are proposed and implemented, however, executives tend to not tune parameters in practical use. As a machine learning techniques, LSTM networks are relatively understandable and easy to mimic the time series forecasting process. Regarding the experiment on 66 products, there are only one fourth products having a relatively low forecasting error, while results of other products are not positive enough, this is because

- The length of raw data is maximum 45 weeks, even less for some products in datasets, which is further divided into training data and test data. It is not sufficient to obtain a robust neural network in this case.

- Seasonality is not considered in the test, moreover, 45 weeks data only contains some short-term seasonality not the long-term seasonality.
- In practical, promotions impacts much on the sales in a way of large fluctuation, which is neither included in the experiment.
- The experiment uses one LSTM network for all products, which means it is not optimized with respect to characteristics of the product.

Without considering the impacts of promotion and seasonality, LSTM network still shows potential for one fourth products in sales forecasting with limited datasets. Further works will be developed in consideration of promotions and seasonality to improve the stability of the network.

Acknowledgements This work was supported by the Research Council of Norway through the Retail Supply Chain 2020.

References

1. Economist Intelligence Unit (2012) Data versus instinct—perfecting global sales performance. http://www.oracle.com/us/products/applications/eiu-oracle-sales-1719684.pdf?sc=EMEAOU13045591MPP008C147
2. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
3. Gonzalez-Dominguez J, Lopez-Moreno I, Sak H, Gonzalez-Rodriguez J, Moreno PJ (2014) Automatic language identification using long short-term memory recurrent neural networks. In: Proceedings of the annual conference of the international speech communication association (INTERSPEECH), pp 2155–2159. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84910072596&partnerID=40&md5=a73a851422179feea8df33e70f2b6af9
4. Bakker B, Linåker, F, Schmidhuber J (2002) Reinforcement learning in partially observable mobile robot domains using unsupervised event extraction. In: IEEE international conference on intelligent robots and systems, pp 938–943. https://doi.org/10.1109/IRDS.2002.1041511
5. Gers FA, Eck D, Schmidhuber J (2001) Applying LSTM to time series predictable through time-window approaches. In: Dorffner G, Bischof H, Hornik K (eds) Artificial neural networks —ICANN 2001: international conference Vienna, Austria, August 21–25, 2001 proceedings, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 669–676. https://link.springer.com/chapter/10.1007/978-1-4471-0219-9_20
6. Wöllmer M, Marchi E, Squartini S, Schuller B (2011) Multi-stream LSTM-HMM decoding and histogram equalization for noise robust keyword spotting. Cogn Neurodyn 5:253. https://dx.doi.org/10.1007%2Fs11571-011-9166-9
7. Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J (2015) LSTM: a search space odyssey. ArXiv e-prints. https://arxiv.org/ct?url=http%3A%2F%2Fdx.doi.org%2F10%252E1109%2FTNNLS%252E2016%252E2582924&v=57f036b2
8. http://colah.github.io/posts/2015-08-Understanding-LSTMs/

A Coarse-to-Fine Matching Method in the Line Laser Scanning System

Leilei Zhuang, Xu Zhang and Weiyan Zhou

Abstract The measurement accuracy of the traditional line laser scanning method is affected by many factors, such as calibration accuracy of the light plane, the detection accuracy of the stripe center, the positioning accuracy of the motion scanning module, etc., resulting in the operation of this method is complex and the precision is not high. In this paper, a line laser binocular stereo vision system is constructed. Because of the binocular stereo vision, it is not necessary to calibrate the light plane and the angle information of the laser rotation has no effect on the measurement accuracy, which avoids the high precision positioning requirement of the motion control module in the traditional line laser scanning method. In order to increase the precision of the stripe center detection, a coarse-to-fine matching algorithm is proposed. First, the extreme pixel of each row is computed in the left image and right image as the rough extraction point for laser stripes center. Second, the foreground and background are extracted and the searching range is limited in a small zone. Third, sub-pixel matching is adopted to refine the laser stripe center correspondence. The proposed method is confirmed in the simulation and experiment. The simulation results show that the RMS of laser stripe center error is less than 0.03 pixel, when the image noise level is less than 7. The three-dimensional measurement experiment shows that when the measurement distance is 1400 mm, the reconstruction precision is as high as 0.314 mm.

Keywords Line laser · Coarse-to-fine · Sub-pixel matching

L. Zhuang (✉) · X. Zhang · W. Zhou
Department of Mechanical Engineering and Automation,
Shanghai University, Shanghai, China
e-mail: zhuangleilei1025@163.com

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_3

1 Introduction

As one of the optical 3D measurement technologies, the line-structured light measurement method has the advantages of simple structure and fast measurement speed. It has been used in many measurement fields, such as the cultural relics reconstruction [1], the defect detection of the product, the parts measurement [2] and so on.

The traditional line-structured light system consists of a CCD camera, a line laser projector and a motion control module [3–5]. The camera captures the scene image after the laser stripe is projected, and the three-dimensional information of laser stripes can be calculated based on the position information between the CCD camera and the laser projector using the triangulation principle. The scanning mechanism is essential to complete the measurement of the entire scene. One scanning way is the relative movement of the entire system and the scene, either the measurement system on the mobile station, or the objects on the conveyor belt. This scanning mechanism requires that the motion control module has a high positioning accuracy, and also demands that the camera shooting and motion module has a high synchronization accuracy, which is complex. Another scanning way is to rotate the laser, keeping the measurement system and the objects intact, and the angle control accuracy of scanning laser in this kind of scanning mode has a great influence on the measurement accuracy [6]. In both scanning methods, the light plane should be calibrated [7, 8]. Therefore, the calibration accuracy of the light plane and the detect accuracy of laser stripe center [9] are key factors affecting the measurement accuracy. In summary, the traditional line laser three-dimensional measurement method is limited by these factors, resulting in its measurement accuracy is not high.

In order to improve the measurement precision of line-structured light method, many researchers put forward many ways in terms of the light plane calibration and the stripe center detection. Hu [10, 11] proposed a light plane calibration method based on the fixed cross ratio, which is low-cost, simple-operation, but low-precision. Lei [12] proposed a stripe center detection method using threshold method and variable orientation template. It computes the laser stripe grey value threshold firstly, and confirms the laser stripe's location based on the adjacent pixels' grey value. This method has a fast computation speed but can only get pixel precision. Hu [13] put forward a stripe center extraction method which can get sub-pixel precision based on the Hessian matrix. It computes the normal of laser line using Hessian matrix, and then gets the laser stripe cross section's first derivative zero crossing using Taylor polynomial expansion, getting the sub-pixel location of the laser center. This method has a high precision but its computation cost is large. Li [14] proposed an algorithm for detecting stripe center combining gradient sharpening with barycenter method, which firstly extracts fringe using gradient sharpening, secondly gets quasi-center of stripe, finally attains the center of stripe with barycenter method. This method is sensitive to noise and complex.

In this paper, a line laser binocular stereo vision system is constructed [15]. Because of the binocular stereo vision, it is not necessary to calibrate the light

plane. The whole scene is measured through the rotation of the laser, while the measurement system and objects remain intact. This system uses the laser stripes as the stereo matching feature, and the angle information of the laser rotation has no effect on the measurement accuracy, which avoids the high precision positioning requirement of the motion control module and the synchronization requirement of the rotating laser and cameras shooting in the traditional line laser three-dimensional measurement method. Based on the suggested system, a coarse-to-fine matching method is proposed for laser stripe center detection. Firstly, the laser stripe is model as a flat laser and the extreme pixel of each row is computed in the left and right image as the rough extraction point for laser stripes center. Second, the foreground, background and width of the laser stripe are extracted and the searching range is limited in a small zone. Third, sub-pixel matching [16, 17] is adopted to refine the laser center correspondence. Last, the scene is reconstructed. The experiment shows that the algorithm is simple, the reconstruction precision is high and the robustness is good.

2 The Framework of the Proposed Method

After the camera calibration, it is necessary to determine the correspondence between the two cameras for the three-dimensional measurement. In this paper, a coarse-to-fine matching algorithm for the line laser binocular stereo vision system is proposed, and it can give the sub-pixel level correspondence, whose algorithm is shown in Fig. 1.

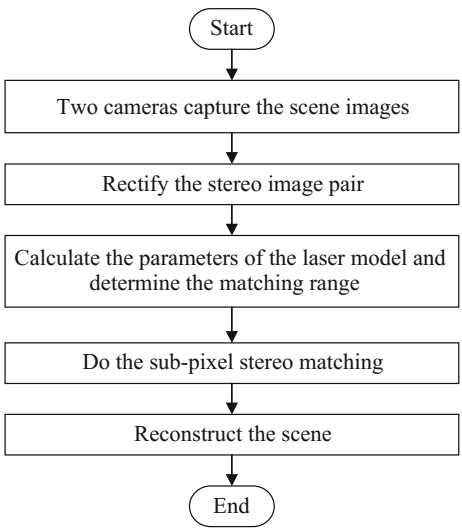


Fig. 1 Three-dimensional measurement algorithm flow chart

First of all, the laser is projected and two cameras capture the scene image which incorporates the laser stripe and scene information. Then, the captured laser stripe images are rectified, thus, the corresponding point in the right image is limited in the current row by the epipolar constraint. After that, the extreme points are computed, and the foreground, background and stripe width are extracted in the left and right image to determine the left image matching range and the right image searching range. Then, the sub-pixel correspondence within the searching range is obtained with the interpolated NCC. Finally, the scene is reconstructed with the triangulation method.

3 Flat Top Laser Model

When the line laser is projected onto the surface of the object, the laser stripe in the image has a certain width. Figure 2 shows the scene image containing the laser stripe collected in the experiment. The laser stripe is modeled as a flat-top laser in this paper, which can be expressed as a rectangular step function with Eq. (1), as shown in Fig. 3.

$$B(x) = \begin{cases} I_1 & x \in [0, x_1) \\ I_2 & x \in [x_1, x_2] \\ I_1 & x \in (x_2, 1] \end{cases} \tag{1}$$

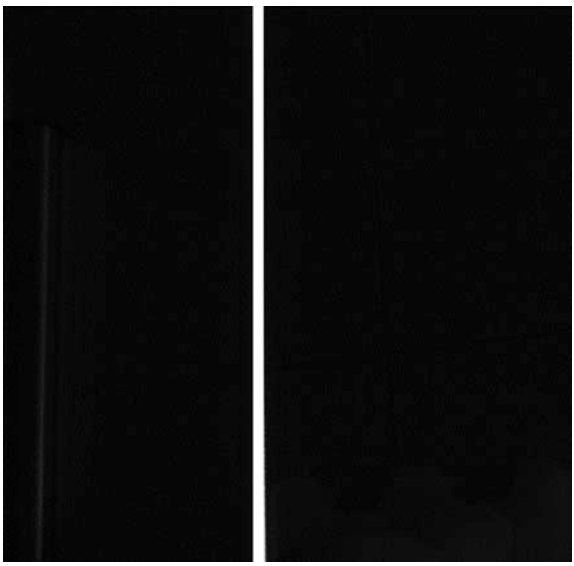


Fig. 2 The image containing a laser stripe

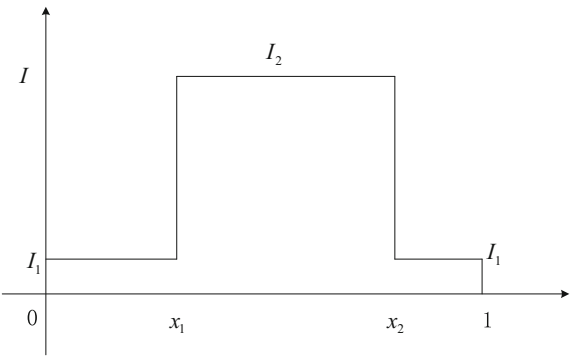


Fig. 3 Line laser cross section model

The center of the laser stripe is $\frac{x_1+x_2}{2}$, and I_1, I_2 are the background and laser stripe brightness respectively. The following is the calculation method of I_1, I_2 and stripe width w.

The 0 order, 1st order, 2nd order and 3rd order moments of the function $B(x)$ can be computed with the Eq. (1), and we get:

$$\begin{aligned}
&\int_0^1 dx = 1 \\
&\int_0^1 B(x)dx = I_1(1 + x_1 - x_2) + I_2(x_2 - x_1) \\
&\int_0^1 B^2(x)dx = I_1^2(1 + x_1 - x_2) + I_2^2(x_2 - x_1) \\
&\int_0^1 B^3(x)dx = I_1^3(1 + x_1 - x_2) + I_2^3(x_2 - x_1)
\end{aligned} \tag{2}$$

Since the image is a digital image, the integral should be transformed into a digital integral, then the 1–3 order moments can be solved by:

$$m_i = \frac{1}{n}\sum_{j=0}^{n-1} (I_j)^i \tag{3}$$

where n is the number of pixels.

According to [18–20], the foreground and background can be solved by:

$$\begin{aligned} I_1 &= \frac{-b-\sqrt{b^2-4ac}}{2a} \\ I_2 &= \frac{-b+\sqrt{b^2-4ac}}{2a} \\ a &= m_2 - m_1^2 \geq 0 \\ b &= m_1 \cdot m_2 - m_3 \leq 0 \\ c &= m_1 \cdot m_3 - m_2^2 \geq 0 \end{aligned} \tag{4}$$

Substitute (4) into the Eq. (2), and we can get:

$$x_2 - x_1 = \frac{m_1 - I_1}{I_2 - I_1} \tag{5}$$

Then, the width of laser stripe w can be calculated through the Eq. (6).

$$w = n(x_2 - x_1) \tag{6}$$

4 A Coarse-to-Fine Matching Method

According to the flat-top laser model in the previous section, we can estimate the coarse extraction point, the foreground, background, and the stripe width of the left and right images, and then the correspondence between the left and right images needs to be determined. In order to improve the accuracy of 3D measurement, this paper uses an interpolated NCC stereo matching method. NCC, that is, normalized cross-correlation [21], is a widely used stereo matching method, and the Eq. (7) is its cost function. The proposed method is insensitive to noise, but only pixel-level accuracy is achieved. In this paper, an appropriate interpolation mechanism is used to obtain the sub-pixel level accuracy on the basis of the NCC matching algorithm. Applying the interpolated NCC matching method to the line laser binocular stereo vision system, a coarse-to-fine matching method is proposed. Figure 4 is the detailed flow chart of the algorithm.

$$ncc(r,c,d) = \frac{1}{(2n+1)^2} \sum_{i=-n}^{n} \sum_{j=-n}^{n} \frac{g_1(r+i,c+j) - m_1(r+i,c+j)}{\sqrt{s_1(r+i,c+j)^2}} \cdot \frac{g_2(r+i,c+j+d) - m_2(r+i,c+j+d)}{\sqrt{s_2(r+i,c+j+d)^2}} \tag{7}$$

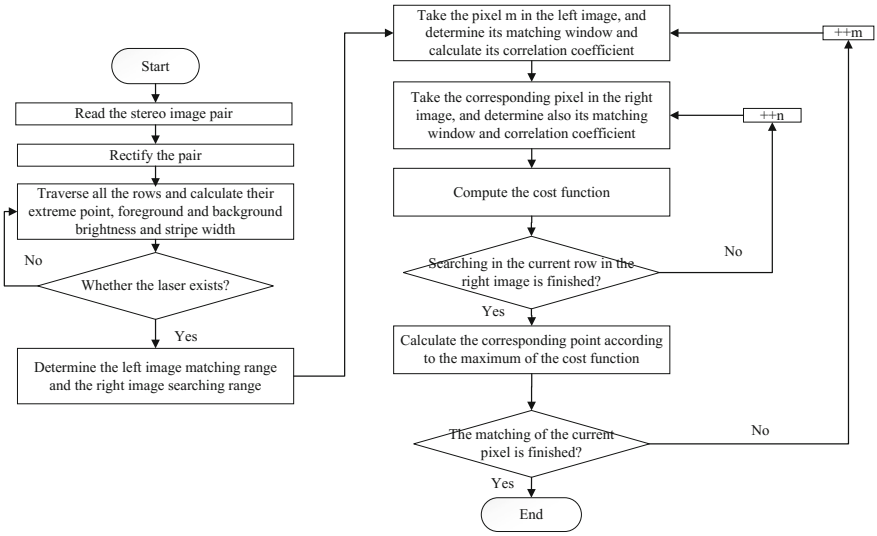


Fig. 4 A coarse-to-fine matching algorithm flow chart

where: g_1, g_2 respectively represent the windows in the first and second image. m_i, g_i $(i = 1, 2)$ respectively denote the mean and standard deviation of all the pixel values in the window of the first and second images, and d is the disparity.

The detailed algorithm is as follows.

Step 1: Read the stereo image pair and rectify them so that they are strictly aligned in the direction of row.

Step 2: Traverse all the rows in the left and right image, calculate their extreme point position colExtreL, colExtreR in the current row, the background I_1, foreground I_2 and stripe width w according to the Eq. (4)–(6).

Step 3: Determine whether there are laser stripes in the left and right images, and the condition of whether the stripe exists is $I_2 - I_1 > T$. T is the threshold decided by the environment.

Step 4: When the current rows in the left and right images have the stripe, the matching range in the current row of left image is set to (colExtreL-2, colExtreL-2), and the searching range in the current row of right image is set to (colExtreR-w, colExtreR + w).

Step 5: Take the m-th pixel in the matching range of the left image, determine its corresponding matching window $W_L(n, m)$, whose size is N1 × N2, whose center is (n, m). Take out the grey value in the window by column and place them in the vector:

$$w(n, m) = [\, w1 \quad w2 \quad \ldots \quad wN \,]^t \tag{8}$$

Define the zero-mean normalized vector $\bar{w}^o(n,m)$:

$$\bar{w}^o(n,m) = \frac{w(n,m) - \bar{w}(n,m)}{\|w(n,m) - \bar{w}(n,m)\|_2} \tag{9}$$

where $\bar{w}(n,m)$ and $\|w(n,m)\|_2$ denote its mean value and Euclidean norm respectively.

Step 6: Take the nth pixel in the searching range of the corresponding row in the right image, determine its matching window $W_R(n, m-d)$ and calculate its correlation coefficient according to the Eq. (10).

$$\rho_{n,m,d} = \bar{w}_L^{ot} \bullet \bar{w}_R^o(n, m-d) \tag{10}$$

The correlation coefficient calculated by the Eq. (10) is the cost function of the NCC algorithm. On this basis, a suitable interpolation mechanism is used to introduce the continuous displacement variable τ between adjacent windows of the right image and define the following N-dimensional vector function:

$$w_R(n, m+\tau) = w_R(n,m) + \tau(w_R(n,m) - w_R(n, m-1)) \tag{11}$$

The cost function of the interpolated NCC matching algorithm is defined by the correlation coefficient and the vector function defined by the Eq. (11):

$$\rho_{n,m,d}(\tau) = \bar{w}_R^{ot}(n, m-d+\tau)\bar{w}_L^o(n,m) \tag{12}$$

Using the definitions of the inner product and zero mean normalized vector of the Eq. (10), and after some mathematical manipulations, the Eq. (12) can be rewritten as

$$\rho d(\tau) = \frac{\rho d + \tau(\rho d - \lambda \rho d - 1)}{\sqrt{(1+\lambda^2 - 2\lambda r)\tau^2 + 2(1-\lambda r)\tau + 1}} \tag{13}$$

where τ is the continuous spatial variable, $\rho(d-1)$ is the correlation coefficient of previous pixel and λ, r are computed like the following.

$$\lambda = \frac{\| wR(n, m-d-1) - \bar{w}(n, m-d-1) \| 2}{\| wR(n, m-d) - \bar{w}(n, m-d) \| 2} \tag{14}$$

$$r = \bar{w}_R^{ot}(n, m-d)\bar{w}_R^o(n, m-1-d) \tag{15}$$

Although the cost function defined in the Eq. (13) is nonlinear with respect to the displacement parameter τ, its maximization results in a closed form solution. We could get its maximum value while $\tau = \tau^0$, and τ^0 is calculated by the Eq. (16).

$$\tau^0 = \frac{\rho d_0 - 1 - r\rho d_0}{\lambda(r\rho d_0 - 1 - \rho d_0) + r\rho d_0 - r\rho d_0 - 1} \tag{16}$$

Step 7: For each reference point in the left image, the interpolation position is determined by the maximum value of the cost function in the right matching image search range. Since the right image matching range is set to [−w, w], it is necessary to match 2w + 1 times for each reference pixel, and the interpolated position $n_{\max} - \tau^0$ corresponding to the maximum value of the cost function is the sub-pixel corresponding point of the left image pixel (where the column coordinates $n_{\max}$ is corresponding to the maximum value of the cost function in the matching image). After getting the correspondence, we can calculate the disparity of the reference pixel.

5 Experiment

In this paper, we use the line laser binocular three-dimensional measurement system shown in Fig. 5 to test the proposed measurement method. The system consists of a 200 nw laser, two PointGrey cameras (the resolution is 1280 × 1024), a stepper motor, a high reflective metal mirror and an industrial computer. Figure 6 is the actual measurement system. In order to verify the stability and reconstruction accuracy of the proposed method, the simulation and the three-dimensional measurement experiment are carried out respectively.

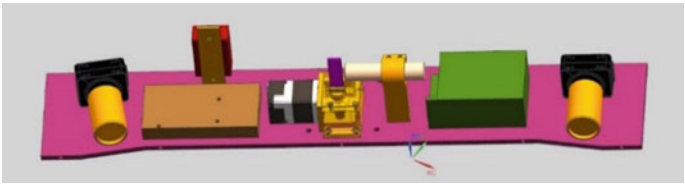


Fig. 5 The line laser binocular measurement system

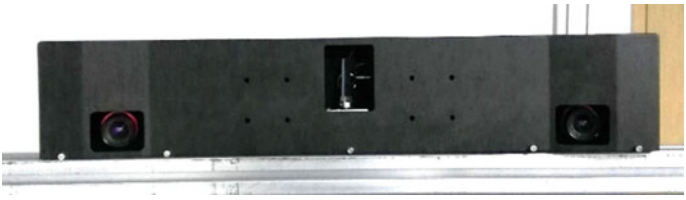


Fig. 6 The actual measurement system

5.1 Simulation

In order to verify the matching accuracy of the proposed method and the robustness of the algorithm, the following simulation experiments are carried out.

First, two theoretical images containing the flat-top laser are designed according to the flat-top laser model of Sect. 3, shown in Figs. 7 and 8. Stripes are vertical, and their width is 11 pixels, whose foreground is 200, and background is 80. One of the images has the stripe center in the 500.25 column, which is the left image of the stereo image pair, and the other stripe center of the image is in the 300th column, which is the right image of the stereo image pair. Use the matching method proposed in Sect. 4 to process the stereo image pair, find all the sub-pixel corresponding points and calculate the disparity of each pixel. Theoretically the disparity of all points is equal to 200.25.

Second, add noise to the theoretical images. Since the noise in the actual camera is close to the normal distribution, the method of adding noise is to generate a set of matrices with two normal distributions, whose means are 0 and standard deviations are 1, 3, 5, 7, 9, 11, 13, 15, 17, and add the matrices to the matching image to get the images with different noise levels.

Then use the method proposed in this paper to calculate the disparity.

Each group of above experiments repeated 300 times, and calculate the maximum value of the error, the RMS and the error mean. The experiment results are shown in Fig. 9.

From the experiment results, we can see that when the noise level is less than 7, the maximum matching error is less than 0.13 pixels, and the error mean is less than 0.05 pixels, indicating that the algorithm has a high matching accuracy. When the noise level is greater than 7 and less than 17, the error maximum is less than 0.23 pixels, and the error mean is less than 0.07 pixels, indicating that when the

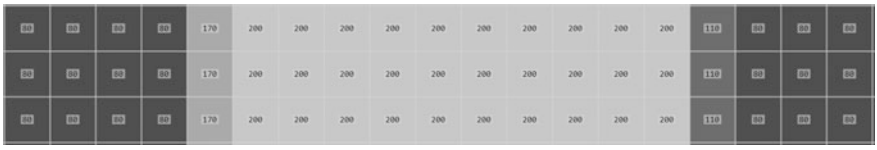


Fig. 7 Left stripe image

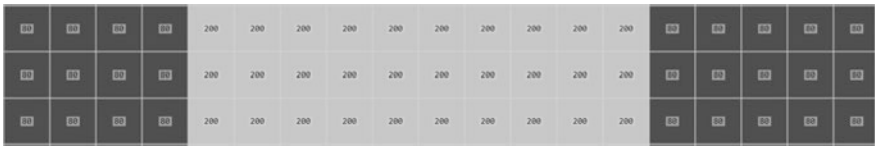


Fig. 8 Right stripe image

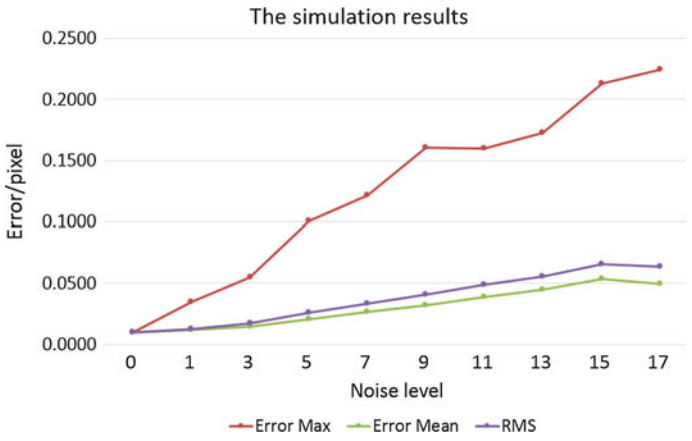


Fig. 9 Simulation experiment results

Fig. 10 The high-precision ball bar

images has relatively large noise, the matching algorithm still has a relatively high matching accuracy. The simulation results show that the matching algorithm used in this paper has a high matching accuracy for the laser stripes, and it is robust.

5.2 The Three-Dimensional Measurement Experiment

In order to analyze the performance of the method in practical application, the following experiments were carried out: reconstruct the high-precision ball bar as shown in Fig. 10, whose center distance is 499.782 mm. The accuracy of the algorithm is determined by comparing the measurements achieved with the recommended method with the standard value.

In the experiment, the measurement distance was 1400 mm, and the field of view of the system is 800 × 600 mm. The mentioned ball bar was reconstructed in 10 different positions which were distributed in most areas within the field of view of the measurement system, including the positions rotating x, y, z axis of camera coordinate system. Figure 11 is an scene image taken in the experiment, which shows the detected laser stripe center in red, and Fig. 12 is a screenshot of the ball bar model reconstructed using the method presented in this paper. The result of the measurement experiment is clearly shown in Fig. 13.

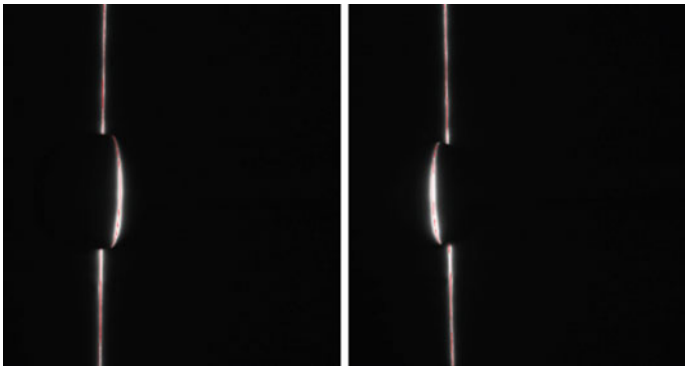


Fig. 11 The scene image

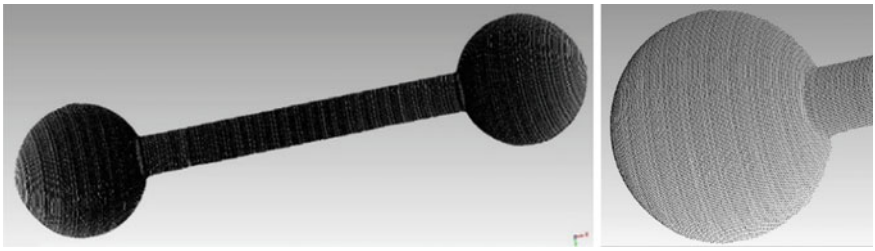


Fig. 12 The reconstructed ball bar model

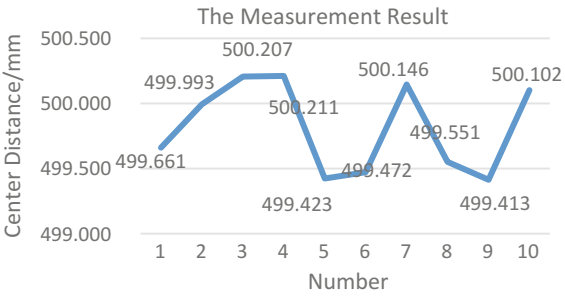


Fig. 13 The error result of three dimension measuring experiments

It can be seen from the Fig. 13 that the maximum error of the center distance of the ball bar is 0.429 mm, and the error mean is 0.314 mm, indicating that this method has a relatively high measurement accuracy at the distance of 1400 mm. Figures 14 and 15 are respectively the workpiece images including the extracted stripe centers and its model reconstructed by this method. Figures 16 and 17 are respectively the scene images including a palm and the reconstructed model.

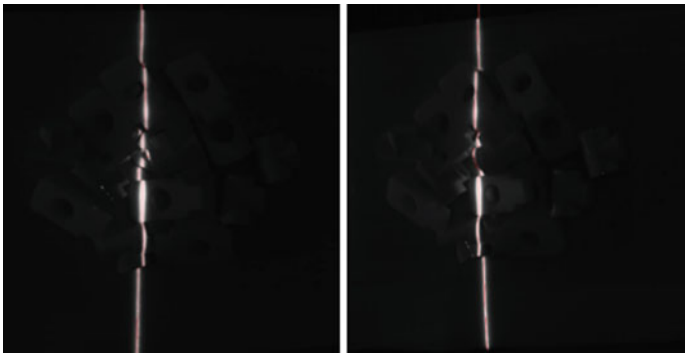


Fig. 14 The workpiece images including stripe centers

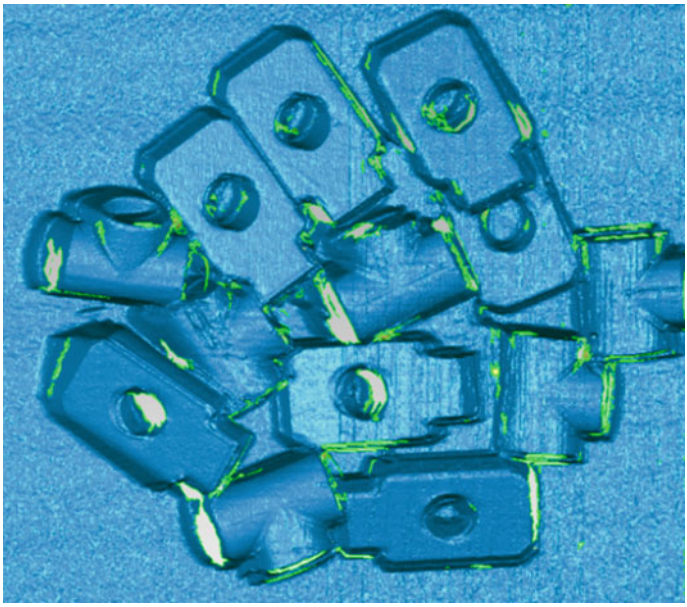


Fig. 15 The reconstructed point cloud of the workpiece

Fig. 16 The palm images including stripe centers

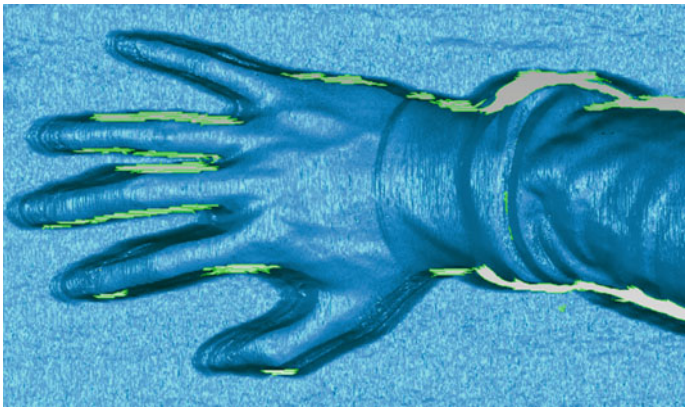


Fig. 17 The reconstructed palm model (detail view)

6 Conclusion

This paper proposes a three-dimensional measurement method combining the traditional line laser technology with stereo matching, which eliminates the process of calibration of the light plane and the process of the stripe center extraction. The experiments show that the whole system structure is simple, low-cost, whose measurement precision is up to 0.314 mm in the distance of 1400 mm, indicating that the method proposed in this paper is suitable for most of the current industrial applications.

Acknowledgements This research was partially supported by the National Nature Science Foundation of China (Grant No. 51575332 and No. 61673252) and The key research project of Ministry of science and technology (Grant No. 2016YFC0302401).

References

1. Zheng J, Guan H, Yi W et al (2012) 3D reconstruction of small-sized cultural relics based on laser scanning and close-range photogrammetry. Int J Digital Content Technol Appl 6 (12):196–205
2. Fu XB, Han JX, Zhang YC et al (2015) Using a line laser to achieve the non-contact dimension measurement of a large thermal forging. Lasers Eng 31(3):237–248
3. Wei H (2008) Research on three-dimensional scanning technology with a single-line laser. Tianjin University
4. Pages J, Salvi J, Garcia R et al (2003) Overview of coded light projection techniques for automatic 3D profiling[C]//IEEE International Conference on Robotics and Automation, 2003. In: Proceedings ICRA, vol 1. IEEE, pp 133–138
5. Xi F, Shu C (1999) CAD-based path planning for 3-D line laser scanning. Comput Aided Des 31(7):473–479
6. Qiu Z, Chen H, Hu W (2013) Embedded angle vision inspection and error compensation for line structured lights. Opt Precision Eng 21(10):2480–2487
7. Duan F (2000) A new accurate method for the calibration of line structured light sensor. Chin J Sci Instrum 21:108–113
8. Zhou F (2004) Field calibration method for line structured light vision sensor. J Mech Eng 40 (6):169–173
9. Wang P (2008) Study on key techniques for automatic 3D structured-light scanning system. Tianjin University
10. Hu Z, Wang Y, Yang J et al (2010) A novel calibration method for three-dimensional modeling system used in virtual maintenance. In: International conference on advanced computer control. IEEE, pp 301–303
11. Chen T, Zhao J, Wu X (2015) New calibration method for line structured light sensor based on planar target. Acta Optica Sinica 35(1):1–9
12. Lei H, Li D, Wang J et al (2003) A method for fast detecting the center of structured light stripe. J Huazhong Univ Sci Technol (Nat Sci Ed) 31(1):74–76
13. Hu K, Zhou F (2005) A fast extraction method for sub-pixel center of structured light stripe. Electro-Opt Technol Appl 20(6):60–63
14. Li Z, Wang C, Shi Y (2008) An algorithm for detecting center of structured light stripe combining gradient sharpening with barycenter method. J Image Graph 13(1):64–68
15. Zhu X, Gao Z (2003) Theoretic research on double-CCD stereoscopic system. Opt Technol 29 (3):298–300
16. Xu Y, Zhou J, Zhou Y (2003) On stereo matching technology. Comput Eng Appl 39(15):1–5
17. Psarakis EZ, Evangelidis GD (2005) An enhanced correlation-based method for stereo correspondence with subpixel accuracy. In: Tenth IEEE international conference on computer vision, vol 1. IEEE, pp 907–912
18. Sun QC, Hou YQ, Tan QC et al (2014) A fast and robust detection algorithm for extraction of the centre of a structured light stripe
19. Sun Q, Hou Y, Tan Q et al (2014) A robust edge detection method with sub-pixel accuracy. Optik-Int J Light Electron Opt 125(14):3449–3453
20. Usamentiaga R, Molleda J, García DF (2012) Fast and robust laser stripe extraction for 3D reconstruction in industrial environments. Mach Vis Appl 23(1):179–196
21. Hanebeck UD (2009) Template matching using fast normalized cross correlation. In: Aerospace/defense sensing, simulation, and controls. International Society for Optics and Photonics, pp 95–102

Pose and Position Measurement in Dynamic Optical Coordinate Measure System

Yong Qi, Xu Zhang, Hao Wu and Wei Cheng

Abstract This paper presents the method of the pose and position measurement in the dynamic optical coordinate measure system. A fixed connected multi-point target and its recognition detection algorithm are designed for the pose measurement. For the position measurement, we designed the coded targets, which can be arbitrarily placed, and its identification detection algorithm. The method proposed in this paper is verified to be correct and reliable by experiments.

Keywords Pose and position measurement · Dynamic optical coordinate measure system · Coded targets · Multi-point target

1 Introduction

With the development of the advanced manufacturing industry, new requirements are put forward for the measurement of 3D geometric dimensions, such as large scale, portable field, digital guide, dynamic real-time tracking [1, 2], etc. The visual measurement task is to measure the workpiece or assembly on the production line so as to make sure they match the preset standard amount. With the dynamic optical coordinate system, visual measure has many obvious advantages, such as non-contact, fast dynamic response, large range, high efficiency, automatic, etc. So it is widely used in the manufacturing and processing industry.

Any object in the space has six degrees of freedom, that is, three directions of translation and rotation around the three directions [3]. Moving object pose and

Y. Qi (✉) · X. Zhang
Department of Mechanical Engineering and Automation, Shanghai University, Shanghai, China
e-mail: qiyong1993@163.com

H. Wu
Robotics Institute, Shanghai Jiao Tong University, Shanghai, China

W. Cheng
Wuxi Research Institute, Huazhong University of Science and Technology, Wuxi, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_4

position measurement is to detect the movement of objects in a particular location of the six degrees of freedom. The study of the pose and position measurement of moving objects is mainly in the field of aircraft control, as well as the observation of satellite pose measurement system. At present it also infiltrates into the field of vehicle automatic driving, intelligent robots and other emerging areas. There are many forms of object pose and position detection methods, which have their own advantages and limitations, and have different application conditions and backgrounds.

The first form is based on GPS Technology. The French SERCEL company produced a differential GPS receiver for pose and position measurement. GPS pose and position measurement system has the advantages of short initialization time, no drift, small size, light weight, low power consumption. But the hardware development is relatively slow, expensive. The location of the pose and position measurement system by the antenna distance and the test environment have greater constraints and lack flexibility. So the pose and position measurement system in the general industry promotion and application has been very limited [4].

The second form is the multi-degree-of-freedom measurement based on the traditional sensors. The more common application of this method is the laser gyro strapdown pose and position measurement system [5].

The third form of the pose and position measurement method is based on the multi-CCD vision measurement technology. This method uses a modern optoelectronic device with the target object or fixed on the target object on a specific target imaging, through the corresponding data processing, to achieve the object of multi-degree of freedom measurement [6, 7]. This method can realize the measurement of dynamic objects in a wide range, but it is necessary to locate the relative position of each optoelectronic device accurately. In this method, a large class is used as a receiver to scan a model of a specific object placed on the measured object to obtain the visual signal of the feature point on the model. After a certain calculation, we can obtain the object of the various degrees of freedom of information [8, 9].

Coded targets can ensure automatic alignment, accurate image point measurement. In recent years, the demand for coded points has greatly dramatically increased [10, 11]. Common coded targets are circularly distributed [12–15] (Fig. 1).

This paper presents a position measurement method based on the coded point, which can realize real-time multi-point tracking. For the pose measurement method,

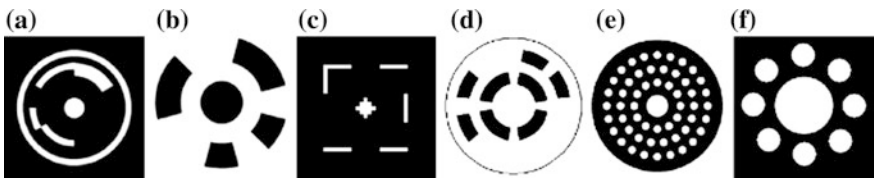


Fig. 1 Different coded targets

this paper presents a novel multi-point target to calculate the pose, which is verified to have strong robustness to noise and high precision by experiments.

2 Principle of Dynamic Optical Coordinate Measure System

In the visual measurement task, far from a comfortable measurement laboratory, portable measuring equipment still has some problems. Production equipment (e.g., cranes, numerical control machine, punchers, machining centers) can generate permanent vibrations. The temperature and humidity are continuously changing. The experience of the operator is also varied. The users of portable measurement solutions will face the daily obstacles that the operators' skill levels are different.

The visual measurement task introduce innovative concepts (e.g., dynamic referencing device, self-identifying) to be tracked continuously through optical cameras. The Fig. 2 shows the operating principle of an optical CMM, which is based on triangulation in Fig. 4. The Fig. 3 shows the principle of the dynamic optical coordinate measure system.

Targets will be identified whether to be a rigid body or not when target triangulation is complete. Thus, for the position measurement in the dynamic optical measurement system, the design and detection of the feature point is an important part. For the pose measurement, the design and detection of the feature target is another important part.

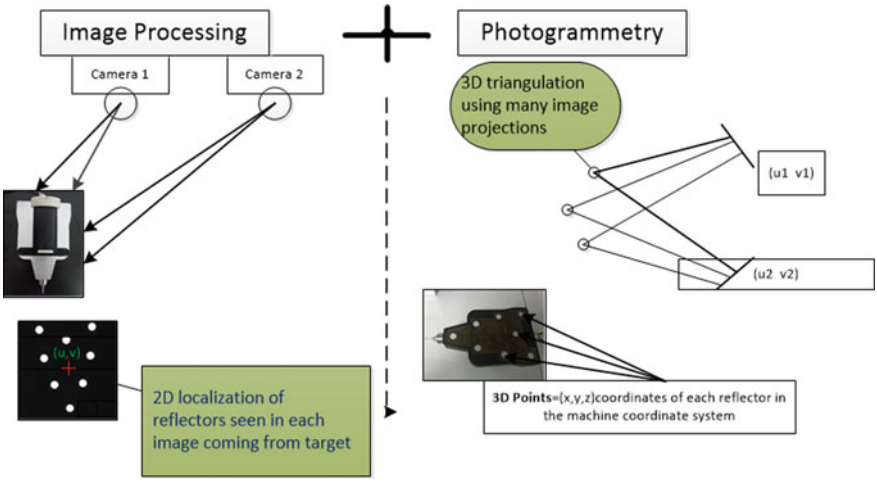


Fig. 2 Operating principle of an optical CMM

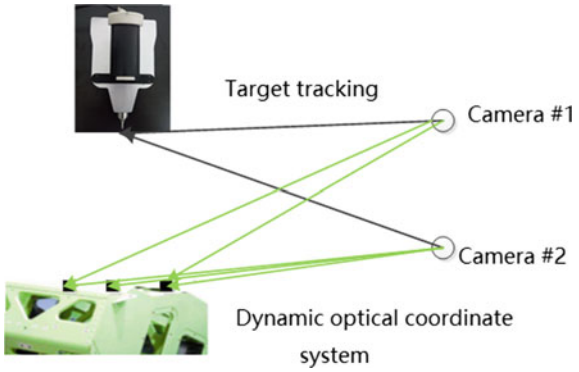


Fig. 3 Dynamic optical coordinate measure system

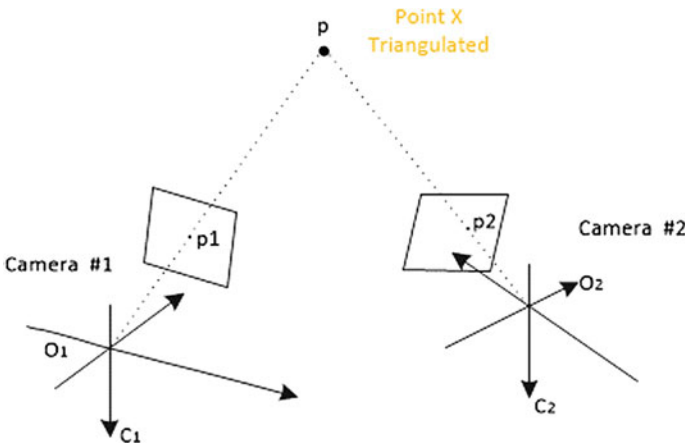


Fig. 4 Principle of triangulation

3 Pose Measurement in Dynamic Optical Coordinate Measure System

Pose measurement can be conducted in many ways. In the visual field, it is common to design targets with different structures. At the same noise level, the accuracy of different target structures for pose estimation is different.

3.1 Design of V-8 Target

As is shown in Fig. 5a, the structure layout of the light pen target is T-7 type [16]. The structure (Fig. 5b) shows the structure layout of the light pen, which is the Metronor's first generation product [17]. The structure layout of the light pen in

Fig. 5d is the common general model. In addition, the rest of the other light pen structure are the existing light pen products on the basis of the possible structures of the idea.

As is shown in Fig. 5e, the V-8 target consists of eight circles with a diameter of the same diameter which have been sorted in order. Each of the three adjacent circular centers constitutes an isosceles triangle, and each of the three forms has an equal waist length.

3.2 Simulation Experiment

We assume a virtual perspective camera in the simulation experiment, which image resolution is 5184 × 3456 pixels and principle point is the image center.

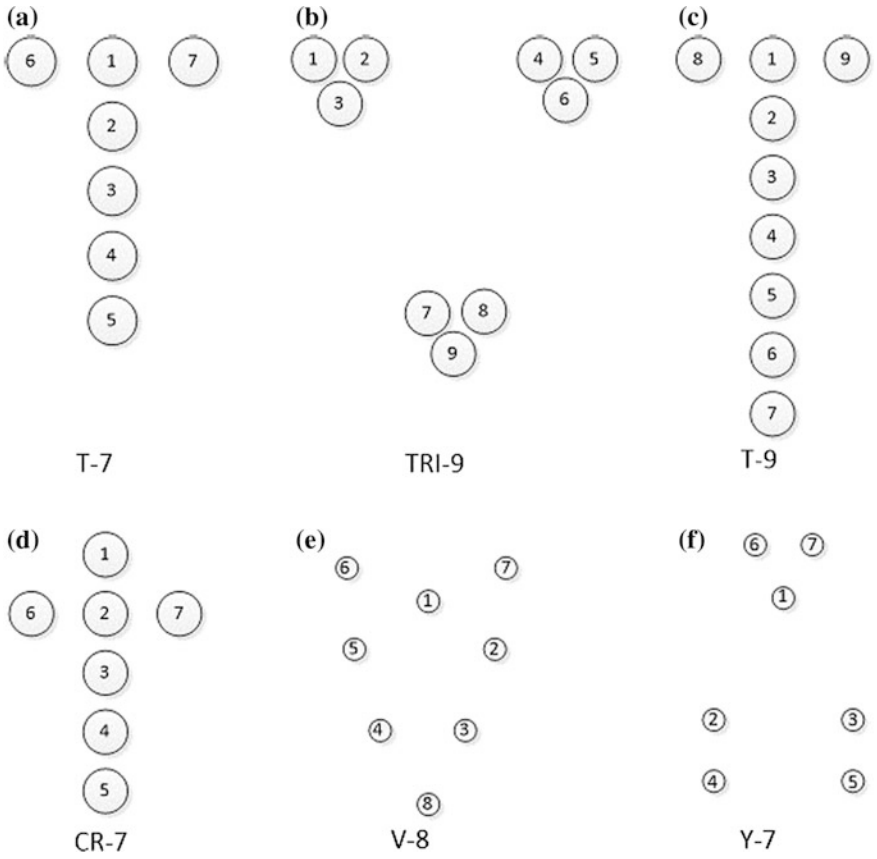


Fig. 5 Different target structures

The 3D reference points in co-planar R_{true} case are generated in the camera frame randomly. Firstly, the z-value is set as zero and these 3D reference points are distributed in the x-, y-range of $[-2, 2] \times [-2, 2]$ randomly. Secondly, we can generate these points randomly by using a rotation and translation matrix. Once R_{true} and the estimated R is measured, the absolute error in degrees will be computed. The absolute error of rotation is defined as $e_{rot}(^{\circ}rees) = \max_{k=1}^{3} a\cos\left(dot(r_{true}^{k}, r^{k})\right) \times 180/\pi$, where r_{true}^{k} and r^{k} represent the kth column of and R respectively. The translation error is computed by the relative difference between t_{true} and t defined as $e_{trans}(\%) = \|t_{true} - t\| \times 100$.

We fix the reference point number and vary the noise deviation level from 0.01 to 0.2 pixels for every different target structure with the DLS method [16, 18] to test the robust characteristic of filter strategy. We run 500 in dependent trails at each noise level. Then we report the average rotation and translation error in the first and second row of the Fig. 6.

As is shown in Fig. 6, it can be concluded from the table that the V-8 target is the best robustness to noise and it has a higher precision than other targets in terms of the accuracy of the pose estimation.

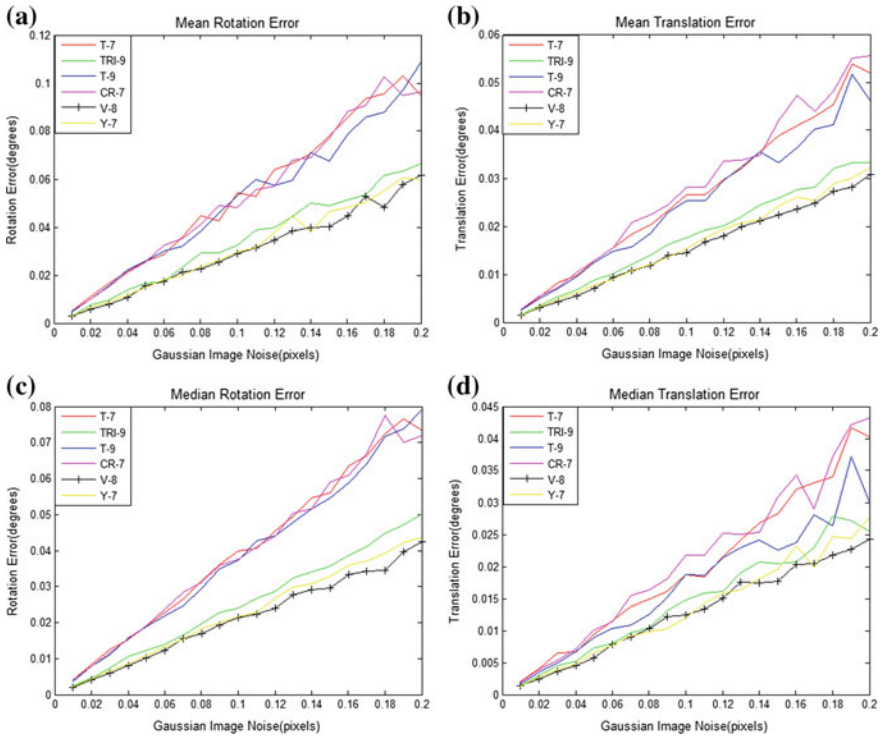


Fig. 6 Compare of six target structures. **a** Mean rotation error. **b** Mean translation error. **c** Median rotation error. **d** Median translation error

3.3 Practical Experiment

In this practical experiment, for purpose of proving the high precision of light pen using V-8 target, the geometric feature is measured, which is known on the standard gauge block with a tolerance level of 0.005 mm. The scene of experiment is shown in Fig. 7. Specially, we measure the Euclidean distance of two points on the gauge block repeatedly. And the nominal value of the distance is 100 mm. In addition, the gauge block is placed at 1.5 m away from the camera. The results are shown in Table 1, where the measured value is showed in the second column and the third column lists the error value. In addition, the standard deviation of error, which is listed in the fourth column, is calculated from the calibration error of the relative position error of the target points. According to the result in Table 1, the accuracy of the V-8 target light pen is very high and it is indicated that the proposed V-8 target is valid and practicable.

3.4 Recognition Algorithm

The recognition algorithm of the V-8 target includes five steps.

Step 1: Find the marked points around which the number of points of equal distance distribution to them is 2, 3 or 4. The results are shown in Table 2.

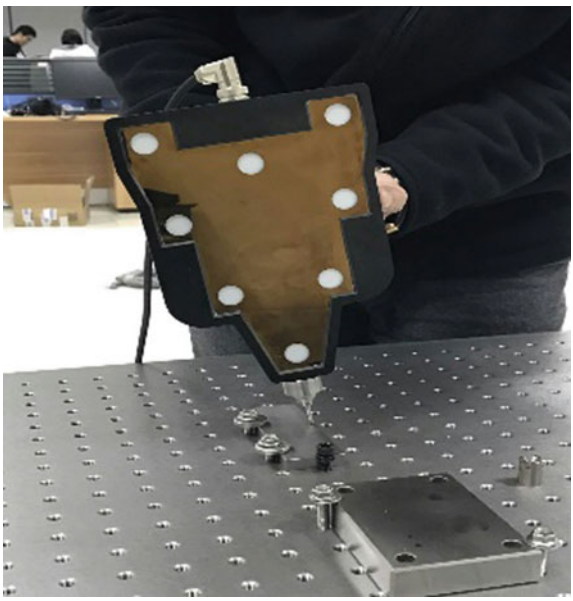


Fig. 7 Measure the actual object

Table 1 Result of real experiment

Serial number	Measured value	Error	Standard deviation error
1	100.009	0.009	0.014
2	100.048	0.048	
3	100.016	0.016	
4	100.002	0.002	
5	100.022	0.022	
6	100.019	0.019	

[a]The unit of Value and Error is millimeter

Table 2 Distance distribution

Number of equidistant distribution	Serial of marked circle			
2	8	4	5	
3	2	3	6	7
4	1			

Step 2: According to the result in Table 2, we can identify the point one easily. Just point one is the point around which distributes four equidistant points to it.
Step 3: Only Point eight is the point around which distributes two equidistant points and it doesn't belong to the points which are distributed around the point one equidistantly. Accordingly, we can easily identify the point eight.
Step 4: We can identify the point four or the point five by the symbol of the dot product results of two vectors between the direction vectors formed from point one to point eight and from point one to point four or two vectors between the direction vectors formed from point one to point five.
Step 5: In a similar way, we can successfully identify other points in turn.

3.5 Complement to Recognition Algorithm

In practice, as is shown in Fig. 8a, because the camera is relatively fixed, and sometimes inevitably the tilt angle of the V-8 target is too large relative to the camera imaging plane, resulting in the picture of the eight points more difficult to be identified exactly. Thus, we can correct the image by correcting it to the parallel view, as is shown in Fig. 8b, so that the eight-point distribution distance is roughly the same as theoretically, so it is easier to be identified accurately.

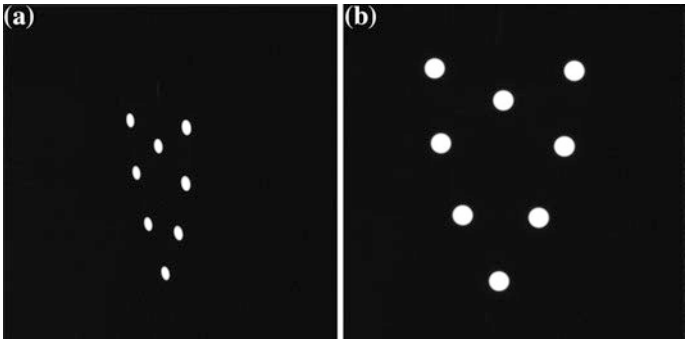


Fig. 8 **a** Uncorrected picture; **b** Corrected picture

4 Position Measurement in Dynamic Optical Coordinate Measure System

In the visual field, coded targets are often used as objects to be tracked and used to establish coordinate system.

4.1 Design of Coded Targets

The coded target is made up of a circle and a ring coded band located around the center. Since the ring code has the characteristics of cyclic reciprocation, the ring coded band of each coded target has no definite starting point. For convenience of use, we select the minimum binary value among the several binary sequences each coded target generates and convert it to a decimal number as a value.

The correspondence between the coded target and the decimal code is shown in Fig. 9. We can find that each coded target is symmetrical distribution to facilitate the identification of the algorithm.

4.2 Recognition Algorithm of Coded Targets

The recognition algorithm of coded targets includes three steps.

Step 1: Detect all the ellipses in the picture and filter out the center circles of the possible coded targets. Since the gradient-based ellipse detection function takes a long time, the original image should be preliminarily detected before using the function. The purpose is to find the possible area of the center circle of the coded target and to narrow the range of process using the ellipse detection function [19].

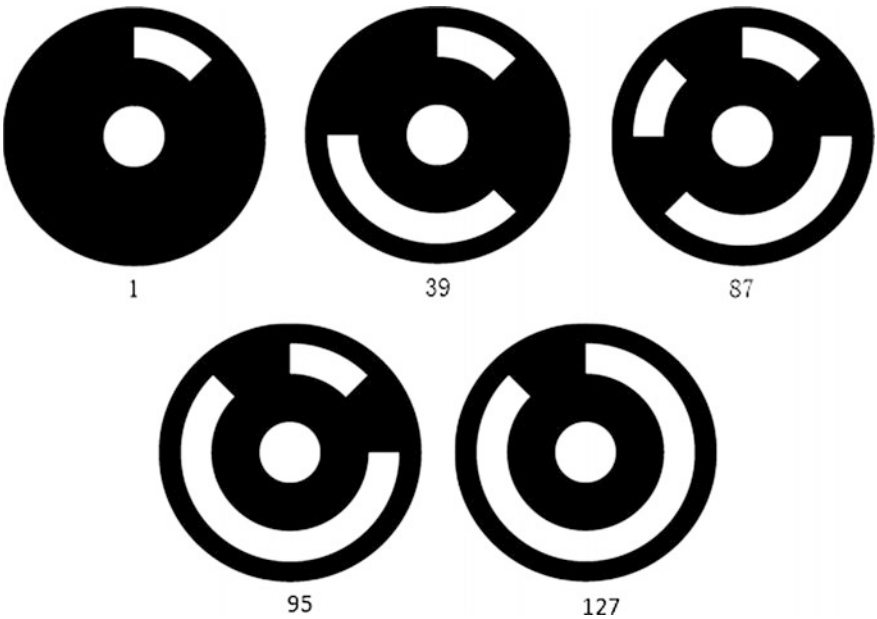


Fig. 9 Coded targets

(a) Load the image and convert it to an 8-bit 1-channel grayscale image.
(b) Use the contour-based ellipse detection function to detect all ellipses in the picture.
(c) Filter the ellipses for the first time to exclude the ellipse with a ratio of major and minor axes greater than 1.6 or with a semi-long axis less than 4 pixels.
(d) Filter the ellipses for the second time. Use a custom function to calculate the distance between the two centers of the ellipse. If the distance between two centers is less than the radius of the larger circle in the two circles, the larger circle will be removed.
(e) Filter the ellipses for the third time to get rid of too large circles. First calculate the mean value of the semi-major axis of the remaining ellipse after the second screening, and then exclude the ellipses that are misdetected according to the degree of deviation from the average value. Since the ellipse with a semi-major axis of less than 4 pixels has been removed, and only the ellipse with too large a long axis is removed here. The purpose of this step is to eliminate the interference of some environmental factors.
(f) Filter the ellipses for the fourth time to eliminate the circle that does not conform to the morphological characteristics of the coded target. According to the design size of the coded target, the connection domain is detected in the area around the ellipse. Exclude the ellipses that do not have a concatenated domain around the center circle or the area of the concatenated domain that is too small.

Step 2: The possible center circles of the coded targets are selected for ellipse detection. The ellipse detection of the target area is carried out by using the gradient-based ellipse detection function [19]. The aim is to finally determine the center circles of all the coded targets and calculate their center coordinates accurately.

(a) A rectangular area around the center of the ellipse after the fourth screen is taken as the target area for the ellipse detection.
(b) Set the parameters in the function to the target area of the previous step and all the target areas are subjected to ellipse detection in turn Load the image and convert it to an 8-bit 1-channel grayscale image.

Step 3: Calculate the code value of each coded target. Follow the steps below to calculate the binary code corresponding to each coded target one by one and convert it to a unique decimal code.

(a) Centered on the ellipse found by the ellipse detection function and intercept the rectangular area around.
(b) Cover the center circle and the peripheral of the ring-coded band with black pixels, binarize the image and find the connected domain.
(c) Find the area of each connected domain and exclude the connected area with too small area. In that case, the remaining connected domain is the ring-coded band.
(d) Judge the number of the ring-coded bands, and if the number of the ring-coded bands is zero, it is not a coded target.
(e) If the number of the ring-coded bands is four, there is only one kind of coded targets in this case, so the binary code of the coded target can be determined.
(f) Also, if the number of the ring-coded bands is three, there is only one kind of coded targets in this case, so the binary code of the coded target can be determined.
(g) If the number of the ring-coded bands is not four or three, then the calculation needs to be continued. We need to use the method of connected domain to calculate the area of the center circle.
(h) Calculate the ratio of the ring-coded band and the center circle, and determine the corresponding number of each ring-coded band according to the size of the design of the coded target.
(i) If the number of the ring-coded band is one, the binary code of the coded target can be determined in the case of knowing the corresponding number of ring-coded bands.
(j) If the number of the ring-coded bands is two, you need to continue the calculation. Calculate the centroid of each ring-coded band using the contour moments, and then calculate the angle of the connection line between the centroid and the center of the circle. According to the angle between the two ring-coded bands, it is determined that they are separated by several units, and finally the binary code of the coded target will be calculated.
(k) Convert the binary code to the decimal code.

5 Application

Our experiment platform is based on the dynamic optical coordinate system. The proposed V-8 target is applied in the novel light pen (Fig. 10a) and we can track the trajectory of the center of the light pen probe and the pose changes of the light pen with the two calibrated cameras at a speed up to 30 frames per second (Fig. 10b). The picture (Fig. 10c) is taken by the camera and the eight points of the target are identified precisely and sorted in order in real time.

The trajectory of the moving light pen probe center is plotted in Fig. 11. The pose changes is showed in Fig. 11. The variable 'dx' represents the amount of rotation around the x-axis. The variable 'dy' represents the amount of rotation around the y-axis. The variable 'dz' represents the amount of rotation around the z-axis (Fig. 12).

Based on the theory of the arbitrarily placed multi-point structure, we deigned five coded targets (Fig. 13) which are easy to be identified and tracked in real time. We can use one coded target or several coded targets stick to the moving object to

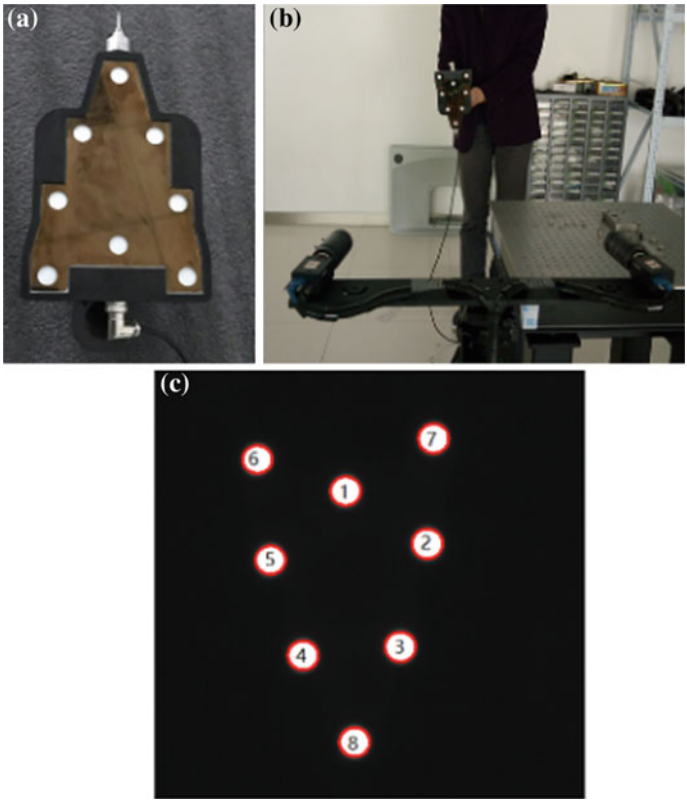


Fig. 10 V-8 target

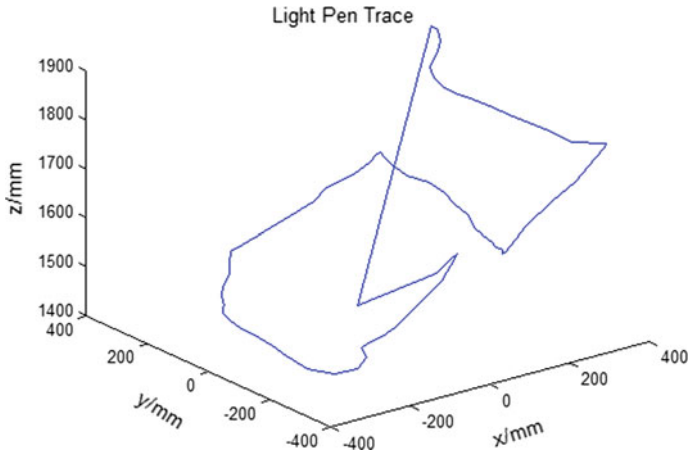


Fig. 11 Motion trajectory of V-8 target

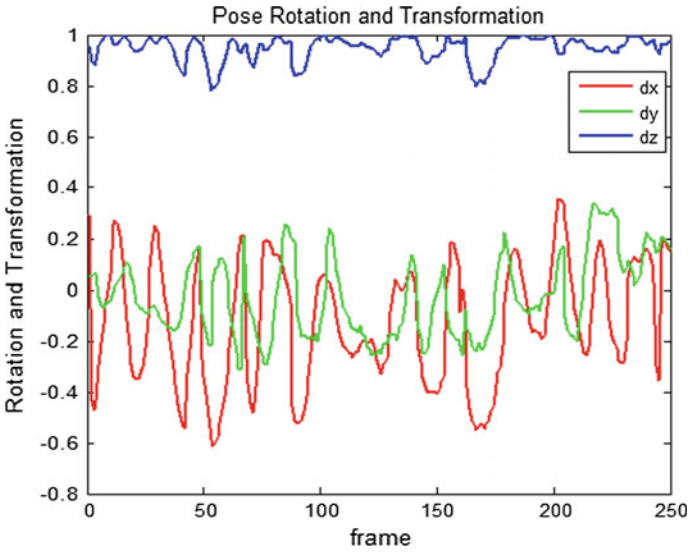


Fig. 12 Pose rotation and transformation

track its motion trajectory including a single point movement and multi points moving together. The Fig. 14a, b are taken by the camera, the coded targets in the picture has been correctly identified and the center point coordinates have been measured in real time. The Fig. 15 shows the trace of a single one coded target and the Fig. 16 draws the trace of several coded targets moving together.

Fig. 13 Coded targets

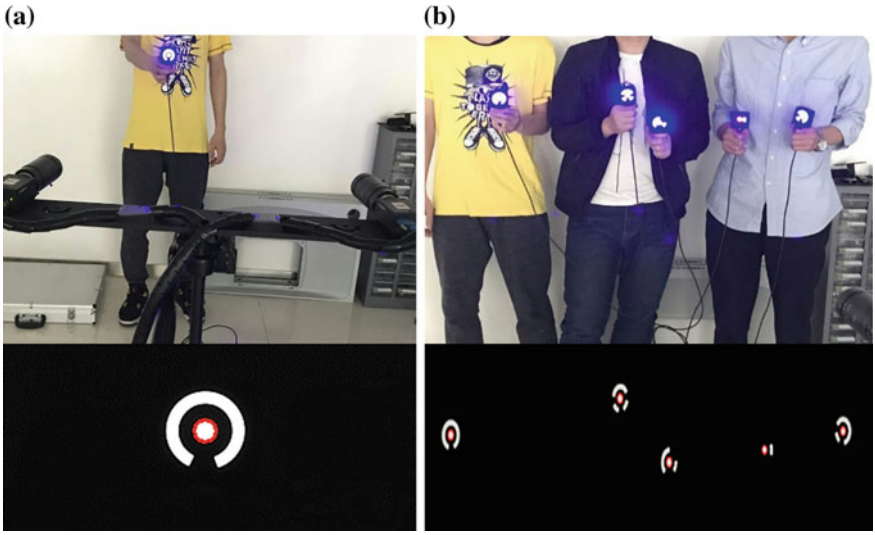


Fig. 14 Movement of different coded targets. **a** Single code moves. **b** Five codes move together

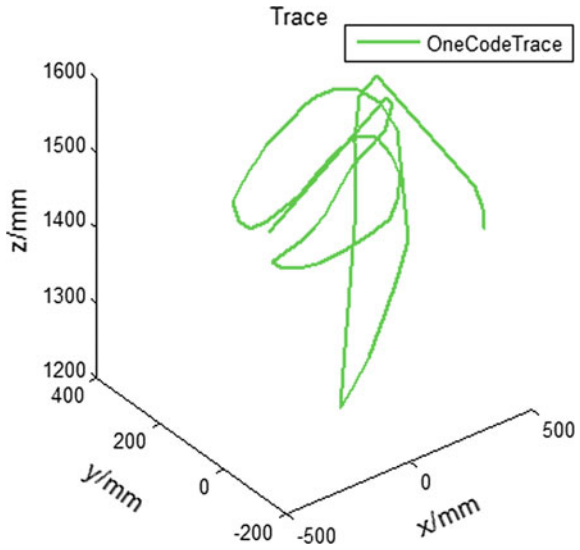


Fig. 15 Trace of a single coded target

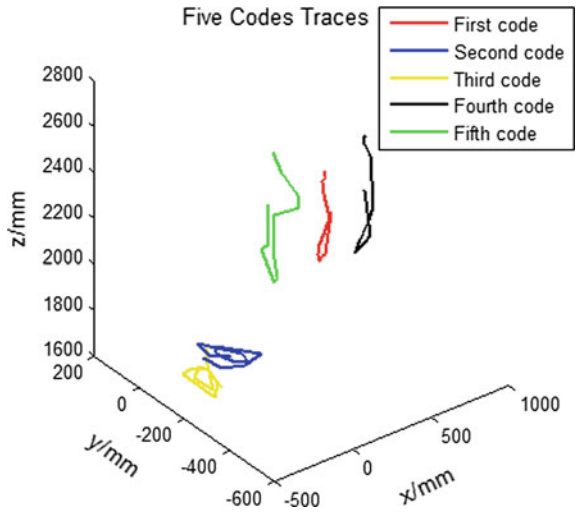


Fig. 16 Trace of five coded targets

6 Conclusion

In this paper, a novel and easily identified V-8 target for the pose measurement in the dynamic optical coordinate system which is applied to the light pen [20] is proposed. It has four stages including the design of the V-8 target, the simulation experiment, the practical experiment and the recognition algorithm. The application of the light pen proves reliable precision and easy operation of the designed V-8 target for object pose measurement.

A set of five coded targets which can be arbitrarily placed are proposed, designed successfully, and easily identified. If we stick the coded targets to the moving object, we can get the motion trajectory in real time. Thus, we can use the coded targets to establish a dynamic coordinate system to improve the measurement accuracy.

Acknowledgements This work was partially supported by the National Natural Science Foundation of China (grant Nos. 51575332, and 61673252), and the National key research and development program (grant No. 2016YFC0302401).

References

1. Larue JF, Viala M, Brown D, Mony C (2012) Dynamic referencing in 3D optical metrology for higher accuracy in shop floor conditions. In: Proceedings of the coordinate metrology systems conference
2. Zhang GJ (2008) Machine vision. Science Press, Beijing, pp 166–173
3. Gao W, Arai Y, Shibuya A, Kiyono S, Hong C (2006) Measurement of multi-degree-of-freedom error motions of a precision linear air-bearing stage. Precision Eng 30(1):96–103
4. Xu JN, Zhu T, Bian HW (2003) Analysis of GPS attitude measurement technology. J Naval Univ Eng 15(3):17–22
5. Liu Y, Wang YT, Hu XM (2002) Study on three-degree-of-freedom localization algorithm for measuring the attitude of moving objects. Comput Meas Control 10(6):363–365
6. Zhang C, Shen ZK, Zhang SQ (2002) Measurement technology based on stereo vision. Syst Eng Electron 24(9):126–129
7. Hu BJ, Zeng L, Xiong W (2007) Study on target attitude measurement based on stereo vision. Comput Meas Control 15(1):27–28
8. Fu QF, Cui YP, Ge XW (2007) Study on three-dimensional attitude measurement method of space axisymmetric target. Syst Microsyst 3:007
9. Tang ZL, Ma CW, Liu B, Dan JL, Chen LH (2004) Study on three-dimensional attitude of space target by single station optical measurement. Acta Photonica Sinica 33(12):1480–1485
10. Russo FA, Knockeart RP (1972) Automated data acquisition with an optical code reader. Bendix Tech J 5:48–52
11. Fraser CS, Brown DC (1986) Industrial photogrammetry: new developments and recent applications. Photogram Rec 12(68):197–217
12. Ahn SJ, Rauh W, Recknagel M (1999) Circular coded landmark for optical 3D-measurement and robot vision. In: 1999 IEEE/RSJ international conference on intelligent robots and systems, IROS'99. Proceedings, vol 2. IEEE, pp 1128–1133
13. Ahn SJ, Schultes M (1997) A new circular coded target for the automation of photogrammetric 3D-surface measurements. Opt 3-D Meas Tech IV 225–234
14. Knyaz VA, Sibiryakov AV (1998) Non contact 3d model reconstruction using coded targets. Image 1:2
15. Shortis MR, Seager JW (2014) A practical target recognition system for close range photogrammetry. Photogram Rec 29(147):337–355
16. Cheng W, Zhang X (2015) An iterative refinement method for pose estimation from planar target. In: 2015 IEEE International conference on robotics and biomimetics (ROBIO). IEEE, pp 2277–2282
17. Metronor products, Internet data, www.metronor.com

18. Hesch JA, Roumeliotis SI (2011) A direct least-squares (DLS) method for PnP. In: 2011 IEEE international conference on computer vision (ICCV). IEEE, pp 383–390
19. Ouellet JN, Hebert P (2007) A simple operator for very precise estimation of ellipses. In: Fourth Canadian conference on computer and robot vision, CRV'07. IEEE, pp 21–28
20. Huang FS (2005) Study on the key technique of single camera 3D coordinate vision measurement system using a light-pen. Tianjin Univ, Tianjin

Design of Pin-Point Gate Injection Mould for Shells of Earplugs

Xian Feng, Min Yang and Min Zou

Abstract The whole design scheme of injection mould for shells of earplugs was determined by analyzing its structure and injection molding technology. In order to adapt to the high product batch and high requirement for appearance quality, an angle ejector of injection mold side core pulling was designed. The layout of a type of pin-point mould with one module and four cavities was determined. The selection of parting surface, the calculation of the cavity number and layout, mold core and cavity design and the design of cooling system were mainly introduced. This paper introduces the structure and working principle of the injection mould for the earplug shell in details. The automatic demoulding of the pouring material is realized. The mold structure is compact which reduces the cost of die manufacturing. It has been proved by practice that the mould is reasonable in structure and the quality of plastic part is stable and has good economic benefits.

Keywords Shells of earplugs · Injection mould · One module and four cavities Parting surface · Pin-point gate

1 Process Analysis of Plastic Part

This plastic part is an earplug shell, and the structure as shown in Fig. 1. The maximum size of the earplug shell plastic part is "28.5 mm × 15.2 mm × 13.1 mm". The average wall thickness of the ear cover, which is connected with the bottom cover of the earphone, is about 1.2 mm, which is inserted into the cylindrical part of the

X. Feng (✉) · M. Yang
School of Electrical and Mechanical Engineering, Taihu University of Wuxi, Wuxi 214064, Jiangsu, People's Republic of China
e-mail: fengxian84@163.com

M. Zou
School of Mechanical Engineering, Changzhou University, Changzhou 213016, Jiangsu, People's Republic of China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_5

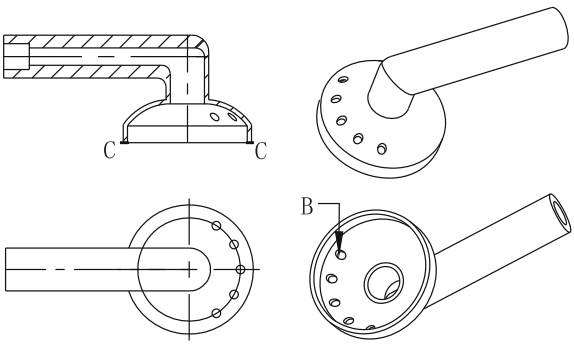


Fig. 1 Earplug shell

earphone line, and the wall thickness is about 2.1 mm. Plastic part is exterior part, the volume is about 0.56 cm^3. The surface roughness requirements are high, requiring no burr, flying edge and shrinkage cavity, must not have demoulding strain, ejector marks, welding marks and other defects [1–3]. Plastic material used ABS, ABS has excellent aesthetic performance and processing performance, suitable for forming a variety of shape, color requirements of the packaging container [4, 5]. This plastic part is the Earplug shell, and the upper cover is assembled with the bottom cover to form a complete earplug. In order to achieve the purpose of assembly with the bottom cover, the inner diameter of the assembly is designed to be ∅ 〖14.8〗 _0^(+0.018) mm [2]. The upper cover has five holes with a diameter of 0.7 mm, and a hole with a diameter of 3 mm is arranged on the lateral side. So it is necessary to use the lateral core pulling mechanism for its lateral processing.

2 Design of Injection Mould Structure

2.1 Design of Parting Surface

The part of the plastic part is shallow dish, and the area is larger, so the main parting surface C-C is set at the lower edge of the maximum diameter of Ø14.8 mm, and because of the rounded corners of the 0.7 mm holes are on the convex surface, so the surface of the hole B should be inside the plastic part, as shown in Fig. 1. Considering the problem of plastic parts to facilitate demolding, and after opening plastic parts should be left on the moving template. At the same time, in order to simplify the mold structure, the slider and the slanting block should also be designed on the dynamic template [6].

2.2 Cavity Number and Layout

There are many types of cavity layout. According to the layout characteristics of the sub-runner, the cavity layout can be divided into two types of balanced and unbalanced layout. According to the layout of the Sub-Runner shape, the cavity layout can be divided into X-shaped layout, I-shaped layout, H-shaped layout and other forms. Considering the characteristics of the plastic part, which requiring large quantities and the size are small. In order to achieve greater production efficiency and economic benefits, so one module and four cavities were determined. The layout of the cavity should adopt the balanced layout of the sub-runner and I-shaped layout, as shown in Fig. 2. This layout allows the plastic to flow through the sub-runner evenly and quickly into the respective gate and into the cavity [6, 7].

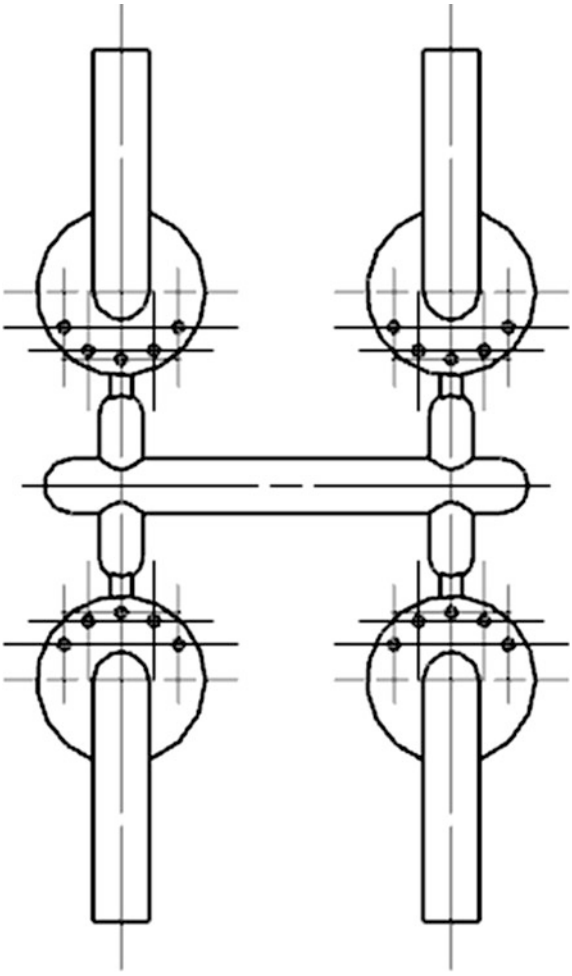


Fig. 2 Sub-runner arrangement

According to this structure, it has advantages of high production efficiency, simple operation, easy release of the flow channel material and high economic efficiency.

2.3 Design of Gating System

Using UG software to analyze the plastic part, and the volume of single earphone is 0.56 cm^3. The injection molding machine selects injection volume of 200 g, and the nozzle diameter is d_o = 3 mm.

2.3.1 Design of Primary Runner

The structure of the primary runner is tapered, and the taper angle is 4 degrees. The length of the primary runner is controlled in 60 mm.

The diameter of the small end which connected with the nozzle of the injection molding machine is:

$$d_1 = d_0 + (0.5 \sim 1)\,\text{mm} \tag{1}$$

According to the formula (1), d_1 = (3.4 ~ 4) mm

Taking d_1 = 4 mm.

The diameter of the big end which connected with the sub-runner is:

$$D = d_1 + 2L\tan(\alpha/2)\,\text{mm} \tag{2}$$

According to the formula (2), D = (7.69 ~ 8.19) mm

Taking D = 8 mm.

D takes 8 mm.

The pit depth is h = (3 ~ 5) mm, the inner wall roughness is R_a 0.4 μm. Considering the material leakage and other issues, it will be the primary runner and the nozzle joint made spherical pit [7–9], as shown in Fig. 3.

2.3.2 Design of Sub-runner

The sub-runner is the feed channel between the main runner and the gate. The contact and friction between the melt and the flow path will occur, and then the melt temperature will be reduced and the flow pressure will be loss. Using the ratio of the cross-sectional area and the perimeter of a flow runner to show the efficiency of the flow runner [1]. Generally, the greater the ratio, the higher the efficiency of the runner. Considering the above two factors, but also consider the sub-runner's volume and surface area, try to ensure that both are relatively small. Taking into account all the above factors, the final decision to cross-section of the cross-sectional shape is designed as a circular structure, as shown in Fig. 2.

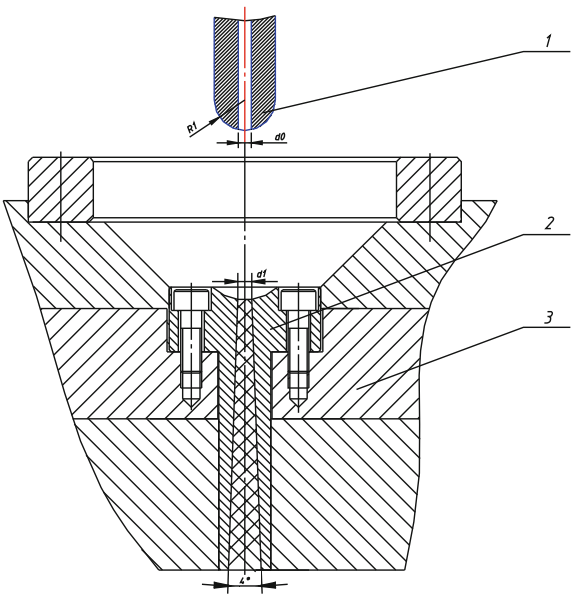


Fig. 3 The shape of primary runner and its relation with the injection nozzle

1-Injector nozzle 2-Sprue bush 3-Fixed die plate

2.4 Design of Shaped Parts

2.4.1 Design of Cavity

The cavity is the external parts of the molding plastic parts, Considering the shape of plastic parts, size and other factors, the size of the plastic parts is smaller, and then choose integral die, the structure as shown in Fig. 4a. This kind of die made by the whole piece of material, and had advantages of good rigidity and high hardness. In addition, the model can avoid the plastic parts producing the phenomenon of joint trace, and the number of the forming parts can be reduced, which is convenient for the processing and assembling of the mould [10].

2.4.2 Design of Core

The core is the inner parts of the molding plastic parts, which have many types of structure. Considering the shape of plastic parts, size and other factors, and then choose integral punch. This type of convex mold structure is reliable, not easy to deformation and no overflow traces, the structure as shown in Fig. 4b.

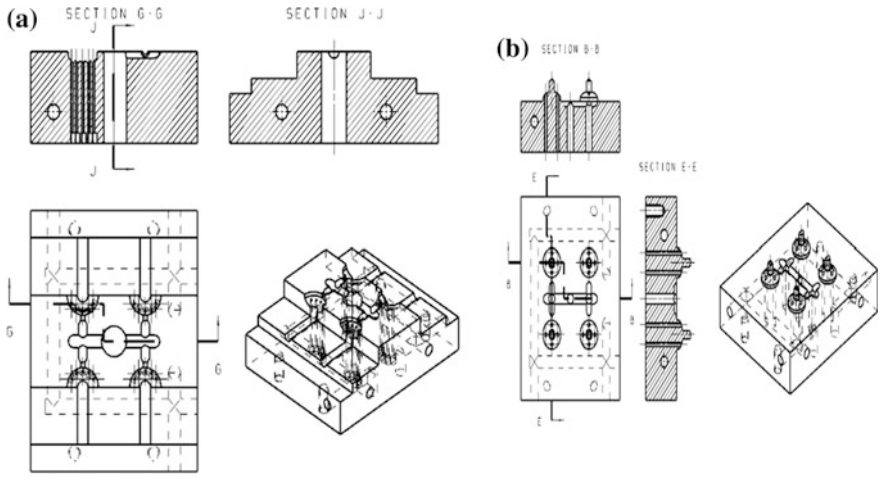


Fig. 4 **a** Integral die. **b** Integral punch

2.5 Design of Cooling System

Temperature control system plays an important role in the design of injection mold, it will directly affect the quality of plastic part and production efficiency. In order to ensure the quality and production efficiency of the plastic part, cooling system is arranged in the mold to control the mold temperature. Because the outside of the plastic parts needs to be pulled out of the core, the core cooling circuit should be designed to avoid interference with the core pulling mechanism, and the design of the 3D cooling water system based on UG can solve the problems. Plastic parts in the cavity of the cooling requirements of uniform, fast, thereby reducing the internal stress of plastic parts, plastic parts to improve production efficiency and quality. Considering the structural characteristics of the plastic parts and the design principles of the cooling system, the cooling system, as shown in Fig. 5, is designed, and using the gap cycle type, taking the diameter of 10 mm and the length of 508 and 538 mm respectively. Figure 5a is core cooling channel(3D), Fig. 5b is cavity cooling channel(3D), Fig. 5c is the 2D and Fig. 5d is cooling system(3D). The cooling liquid is separated into the mould core from the outside of the template, and then flows out from the reverse phase after the double way circulation. Mold and template contact with sealing ring, to prevent leakage. The cooling system has the advantages of simple and reliable structure, fully meet the design principles of the cooling system, and can effectively improve the quality of plastic part, also can improve production efficiency.

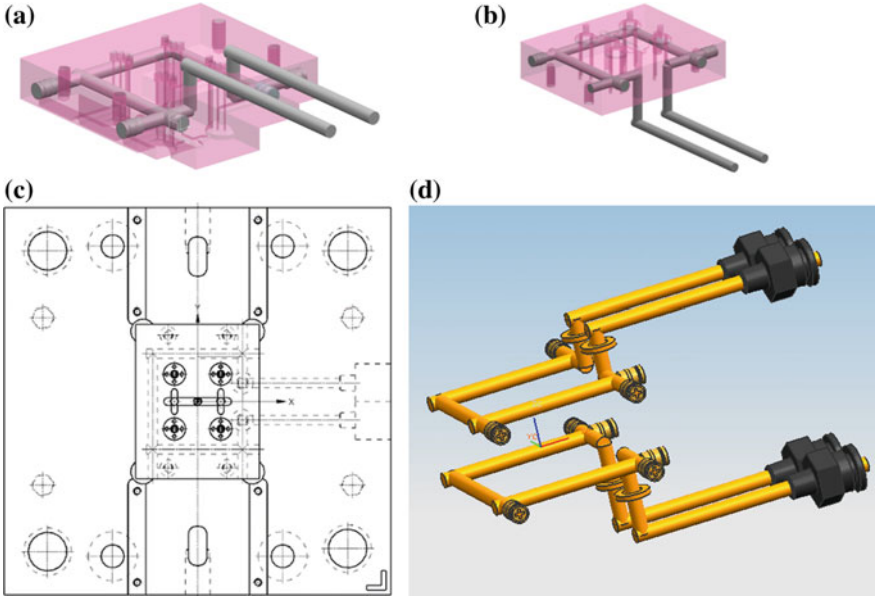
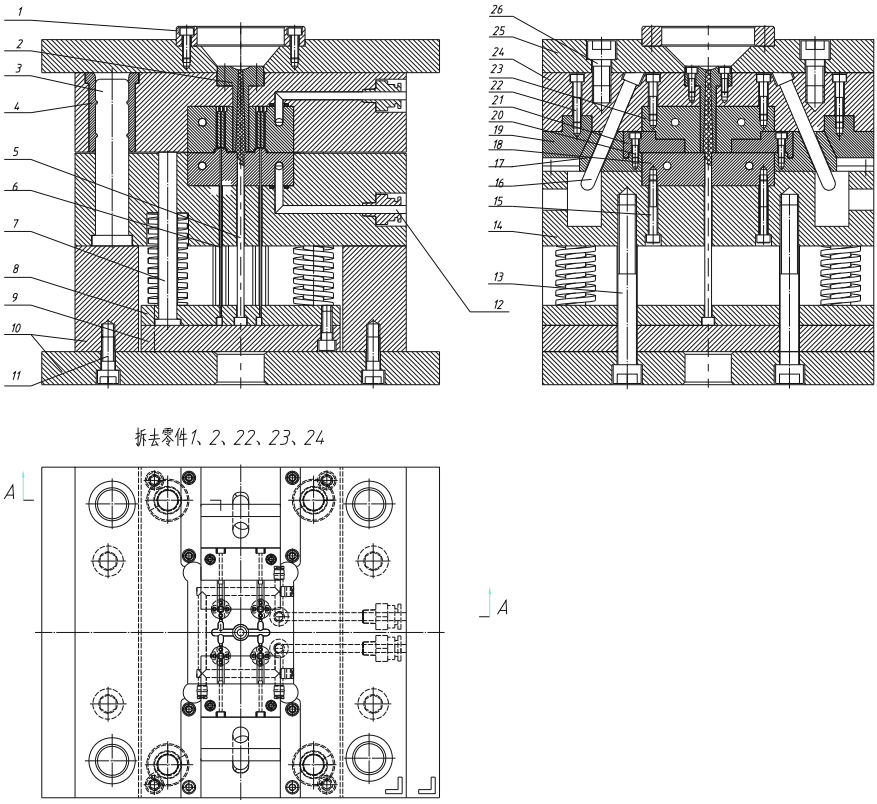


Fig. 5 **a** Cavity cooling channel layout(3D). **b** Core cooling channel layout(3D). **c** Cavity cooling channel layout(2D). **d** Cooling system(3D)

3 Structure of Injection Molds

The structure of the injection mould is shown in Fig. 6. Mold injection molding is mainly composed of injection pressure, opening mould after cooling and solidification, plastic parts introduction and closing. Its working principle is: During injection molding, the mold is in the closed state. The shape and size of plastic parts are formed by cavity 23 and core 18, and the closed precision is guaranteed by the guide pin 3 and the guide sleeve 4. The melt is injected into the mold cavity by the injection molding machine. In this process, it must be ensured that the dynamic template 14 and the fixed template 24 is not melt pressure expansion. After cooling and solidification, under the action of opening force, the dynamic template 14 and fixed template 24 are disposed on the parting surface, and then the plastic part is pulled out from the fixed template with the gating system. The angle pin 16 is driven by the opening force and then acts on the slider 21, forcing the slider to drive the side core to move outwards (4 mm) in the guide groove of the dynamic template 14, realizing the core pulling action. The plunger of the injection molding machine is launched, and then the push rod fixed plate 8 drives the push rod 6 to complete the plastic ejection, completes demoulding. When molding, the return pin 7 is driven by the spring to drive the pushing mechanism and the push rod 6 to reset, and then the injection molding machine drive the dynamic template moves along the clamping direction, at the same time the slider 21 moves inward 4 mm. Finally,



1—Locating ring；2—Sprue bush ；3—Guide pin；4—Guide sleeve；5— Pull rod；6—Push rod；7—Return pin；8—Push rod fixed plate；9—Push rod；10— Moving clamping plate ；11—Screw M10×25；12—Cooling system；13—Screw M6×121；14—Dynamic template；15—Screw M6×42；16—Angle pin；17—Slide block；18—Core；19—Slide locking block；20—Screw M5×14；21—Slider；22—Screw M6×31；23—Cavity；24—Fixed template；25—Top clamping plate；26—Screw M16×25

Fig. 6 Structure of injection molds

the locking block of the slide block19 is used to lock the mold to complete the clamping action, and prepared for the next cycle of injection molding.

4 Summary

According to the requirement of mass production, the layout of a type of pin-point mould with one module and four cavities were determined. Through the software, the plastic parts are designed and simulated in many ways, which improves the

success rate of the mould. It has been proved that the structure design is reasonable, and the precision of the plastic parts is good, which greatly improves the production efficiency and meets the requirements of the enterprise.

References

1. Shen YJ, Liu HY (2016) Mouse bottom shell injection mold design based on PRO/E. Eng Plast Appl 4:82–85
2. Yu SR, Zeng LY, Han W, Wen H (2016) Design of pin-point gate injection mould for covers of earplugs. Eng Plast Appl (8):78–82
3. Guo CG, Peng C, Li Y (2016) Injection mold design of motor canteen rear covers with threads. China Plast (30):109–112
4. Yu ZS (2012) Advances in research on ABS resin. Polym Bull 5:40–46
5. Cao YL (2016) Design of injection mold for powder box top-cover with internal and external multidirections core-pulling. Eng Plast Appl (12):76–79
6. Qu HC (2011) Plastics molding technology and mould design. High Education Press, Beijing
7. Qi XL (2013) Plastics molding technology and mould design. China Machine Press, Beijing
8. Chen ZG (2009) Plastic mould design. Machinery Industry Press, Beijing
9. Li RC (2015) CAE analysis of tail light socket injection and mold UG design. Plastics (6):93–96
10. Huang H (2003) Plastic molding and die. Chemical Industry Press, Beijing

Research on the Principle of Automatic Correcting Machine

Haoyan Wang and Ziqiang Zhou

Abstract In this paper, a set of correcting principle prototype is designed according to the characteristics of the bending deformation and the correcting method of the spool. In order to improve the correcting accuracy of the principle prototype, the prediction model of the correcting stroke is established based on the T-S fuzzy neural network. The results of this model are compared with those of BP neural network. The results show that the model established by T-S fuzzy neural network has faster convergence speed and stronger generalization ability. In order to realize the automation of correcting, the design scheme of automatic correcting machine is put forward based on the principle prototype.

Keywords Automatic correcting machine · Spool · T-S fuzzy neural network

1 Introduction

Resources, environment and energy, as the theme of the development of the society, have been highly valued by lots of countries for many years. Remanufacturing engineering is of great significance to alleviate the shortage of resources and protect the environment through the purchase, dismantling, repair, reuse of waste products and the utilization of materials [1].

As a tool for winding metal wires or steel ropes, the spool is widely used in the metal products industry. But in the case of mechanical processing and even heat treatment has been completed, the bending deformation of the spool often occurs,

H. Wang
School of Mechatronic and Electrical Engineering, China University of Mining and Technology, Xuzhou, China

Z. Zhou (✉)
Jiangsu Key Laboratory of Recycling and Reusing Technology for Mechanical and Electronic Products, Changshu, Jiangsu, China
e-mail: zzq_hefei@163.com

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_6

which results in the stability of the operation of the spool. The density and flatness of the wire wound will also be affected [2]. If we don't correct the bending of the spool, we may cause the waste of the resources.

At present, the workers mainly rely on the experience to determine the correcting stroke of the wheel, this traditional approach is low efficiency and the accuracy can't be guaranteed. Therefore, the research and development of the automatic correcting machine is very important, and the precise prediction of the correcting stroke plays an important role in the correcting accuracy of the automatic correcting machine.

The methods which are frequently used to predict the correcting stroke include the theory of elastic-plastic [3], FEM [4] and empirical formula [5]. Most of the elastic-plastic theories only derive the formula for the regular section, in the derivation process, the model is simplified and assumed, so the accuracy is not accurate. The finite element method is suitable for all kinds of parts, but the accuracy of the method depends on the accuracy of the modeling and input parameters. Although the accuracy of the empirical formula method is higher, different parts need to establish different formulas, so the scope of application is limited.

Due to the complexity of the bending form and the bending deformation is influenced by the nonlinear factors, such as the material, the size, the heat treatment process and so on, these methods have some limitations in the establishment of the correcting stroke model. With the development of intelligent technology, fuzzy control and neural network have a good effect in dealing with nonlinear and uncertain problems.

Therefore, this paper combines fuzzy control with neural network to establish the T-S fuzzy neural network to predict the correcting stroke of spool. Based on the successful experiment of the principle prototype, the design scheme of the automatic correcting machine is proposed.

2 Method for Correcting the Spool

Through the analysis of the spool, it is found that the maximum bending of the spool mainly occurs at the edge of it. First, the displacement sensor is used to detect the deformation of the spool. Then using the electric cylinder to apply a load to the deformed position, as shown in Fig. 1. It produces a plastic deformation which is opposite to the error direction. After unloading, when the residual deformation of plastic deformation is equal to the amount of error, the spool can be corrected [6].

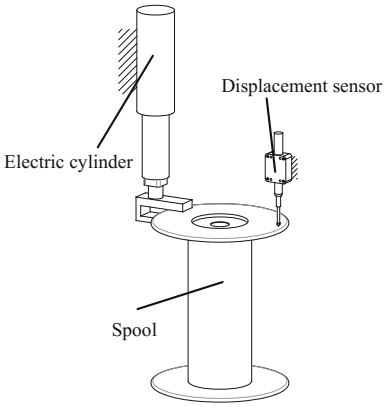


Fig. 1 Method for correcting the spool

3 Bending Correcting and T-S Fuzzy Neural Network for Predicting Correcting Stroke

Through the analysis of the characteristics of the bending deformation and the correcting method of the spool, a set of correcting principle prototype is designed and the correcting experiment is carried out on the spool. In order to improve the correcting accuracy of the principle prototype, the T-S fuzzy neural network model is used to predict the correcting stroke.

T-S fuzzy neural network is the combination of fuzzy system and neural network, The fuzzy reasoning is done by the structure of neural network, the fuzzy membership function and fuzzy rule are improved by learning the sample. It not only has the advantage of fuzzy logic to express human knowledge but also has the advantages of distributed information storage, parallel processing and self-learning ability of neural network [7–9]. In this paper, the T-S fuzzy neural network with double input and single output is designed, which takes the spool's bending deformation and variation as input, and takes correcting stroke as output.

Using the principle prototype to catty out the experiment on the spool to obtain 55 sets of experimental data. The 55 sets of experimental data are used as the sample data to participate in the training and test of the model. 45 sets of data are used as the learning samples of the model, and the 10 sets of data are used to test the generalization ability of the model.

In order to verify the merits and demerits of T-S fuzzy neural network in the prediction of correcting stroke of the spool, compare it with the BP neural network, using the relative error and absolute error of the predicted and true values as the performance index of the evaluation model.

It can be seen from Table 1 that although the prediction results of T-S fuzzy neural network have large errors at the fifth prediction point, the prediction accuracy of other points is better than that of BP neural network. The average relative error of the two models is 4.17 and 6.44% respectively. It can also be seen from Fig. 2, compared with the T-S fuzzy neural network, the BP neural network has a

Table 1 Comparison of two models

Experimental no.	True value/ (mm)	T-S fuzzy neural network			BP neural network		
		Predicted value/ (mm)	Absolute error/ (mm)	Relative error/%	Predicted value/ (mm)	Absolute error/ (mm)	Relative error/%
1	11.43	10.79	−0.64	−5.59	10.68	−0.66	−0.58
2	9.00	8.69	−0.31	−3.44	8.51	−0.49	−5.44
3	12.4	12.37	−0.03	−0.24	12.14	−0.26	−2.09
4	10.9	10.55	−0.35	−3.21	10.04	−0.86	−7.89
5	9.2	7.16	−2.04	−22.17	8.82	−0.38	−4.13
6	7.1	7.35	0.25	3.52	8.47	1.37	19.29
7	13.5	13.7	0.2	1.48	13.17	−0.33	−2.44
8	10.8	10.66	−0.14	−1.29	9.19	−1.61	−14.91
9	12.7	12.61	−0.09	−0.709	12.14	−0.53	−4.17
10	8.2	8.08	−0.12	−0.15	8.48	0.28	3.41
Maximum error		−2.04			−1.61		
Average relative error		4.17%			6.44%		

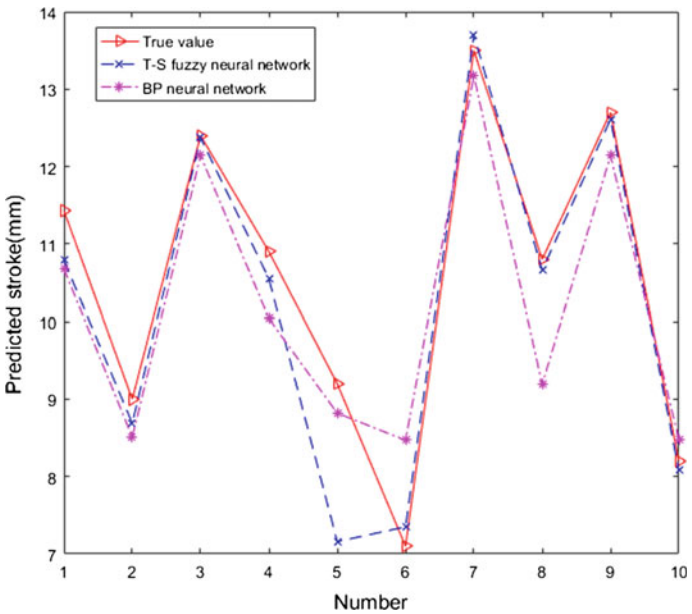


Fig. 2 Comparison of forecasting results

larger deviation between the predicted value and the real value in many place. The fitting effect between the predicted value and the true value of T-S fuzzy neural network is better.

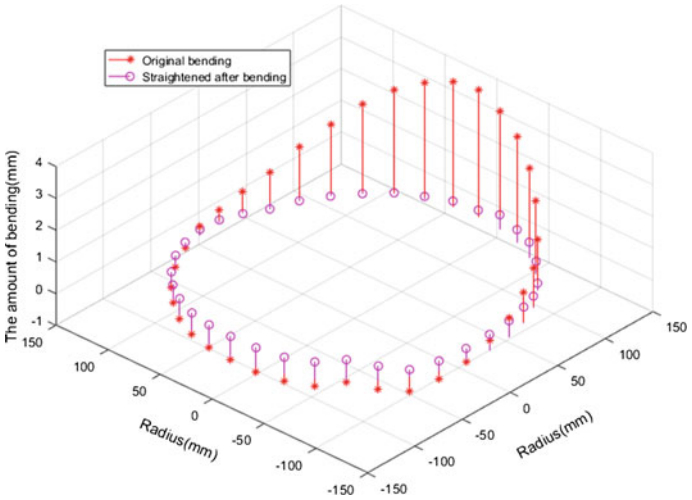


Fig. 3 Bending contrast

In the process of training, the T-S fuzzy neural network can only meet the convergence requirements after 30 training, while the BP neural network tends to converge after the 100 training.

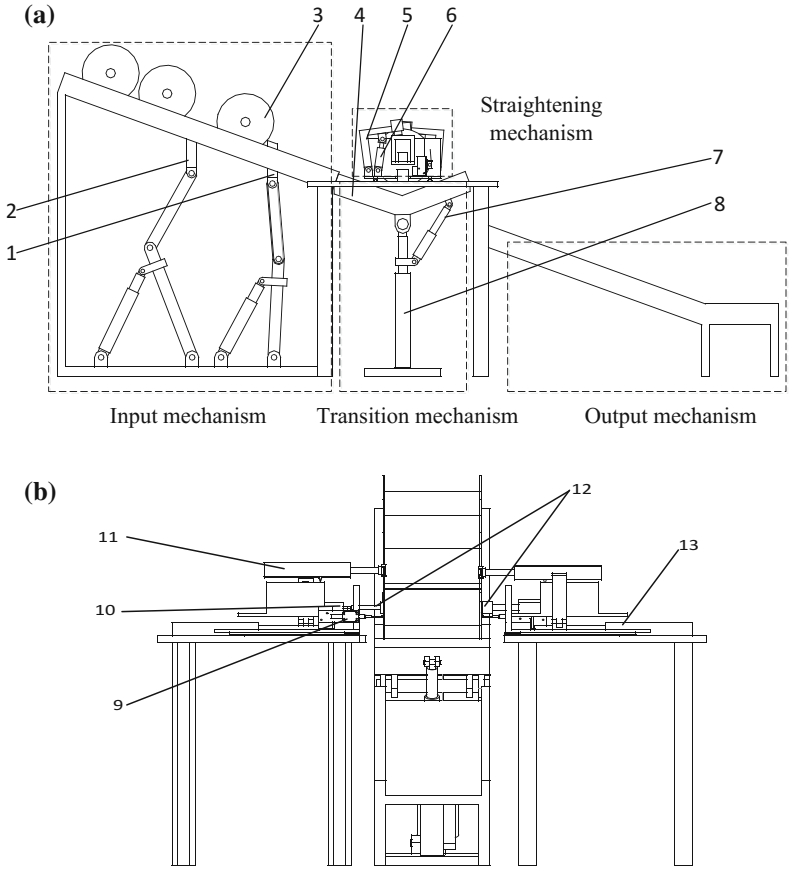
Figure 3 shows the bending deformation of spool which is corrected and not corrected. After a correcting, the bending amount of each point reaches the required range of [−1mm, 1 mm].

4 The Design of Automatic Correcting Machine

From the above experiment, we can see the validity of the principle prototype and the feasibility of the correcting model. In order to improve the efficiency and precision of the correcting. We design an automatic correcting machine based on the principle prototype. In this machine, the conveying, clamping, testing, correcting and output of the spool are completed under the control of the computer, the degree of automation is high. According to the role of each part, it can be roughly divided into: input mechanism, transition mechanism, correcting mechanism and output mechanism, as shown in Fig. 4.

The correcting process can be divided into the following five steps:

1. The conveying mechanism sends the spool which should be corrected to the support plate in the transition mechanism by controlling the movements of the baffles 1, 2.
2. The transition mechanism adjusts the height and angle of the support plate by driving the hydraulic cylinders 2, 3, so that the axis of the spool is aligned with



1:baffle 1 2:baffle 2 3:spool 4:support plate 5:L-shaped bar 6: hydraulic cylinders 1 7: hydraulic cylinders 2 8: hydraulic cylinders 3 9:displacement sensor 10:servo motor 11: electric cylinder 12:centre 13: hydraulic cylinders 4

Fig. 4 Automatic correcting machine

the centre in the correcting mechanism. Drive the hydraulic cylinder 4 to clamp the spool.

3. The servo motor drive the centre to make the spool turn a circle, meanwhile, the displacement sensor detects the amount of bending of the spool and transmits the date to the IPC, the data is analyzed and processed by the IPC to get the correcting position and the correcting stroke, then drive the motor to turn the spool to the corrected position.

4. The hydraulic cylinder 1 drives the L-shaped bar to rotate, then L-shaped bar drives the electric cylinder into the corrected position and corrects the spool.
5. After the correcting is completed, check the amount of bending of the spool, and judge whether it needs to be corrected again according to the requirements.

5 Conclusion

In this paper, we use the self-made principle prototype to do the correcting experiment on the spool and apply the T-S fuzzy neural network to the predict the correcting stroke of the spool. The experimental results show the validity of the principle prototype and the feasibility of the correcting model. The design scheme of automatic correcting machine which was based on the principle prototype can greatly improve the efficiency and precision of the spool.

Acknowledgements This work was financially supported by Key projects of natural science research in universities and colleges in Jiangsu Province (No. 15KJA460001); Open Research Fund by Jiangsu Key Laboratory of Recycling and Reuse Technology for Mechanical and Electronic Products (No. KF1505)

References

1. Zhou Z, Dai Gg (2016) A study of technical architecture of intelligent remanufacturing. J Changshu Inst Technol 30(2):1–3
2. Wang P, Liang J (2015) Modeling and simulation analysis of spooler winding process. Comput Simul 32(11):238–244
3. Chen H, Xiong G, Li J (2005) Study on a method for the determination of correcting strock based on F-δ model and its application. Mach Des Res 21(4):70–73
4. Li J, ouU H, Xiong G, Wang X (2005) Research on theoretical model of press correcting process and its experiment. J Mech Strength 27(5):636–639
5. Zhang X (2009) Research on prediction of correcting stroke and modeling of precision linear guide rail. Wuhan: Master's thesis of Wuhan University of Technology
6. Zhao L, Shen Z, YE J, Jiang W (2007) Research on intelligent correcting processing technology for axis parts. Modular Mach Tool Autom Manuf Tech (12):12–16
7. Li G, Yang L (2013) Neural fuzzy predictive control and its MATLAB implementation. Publishing House of Electronics Industry
8. Sun Y, Yan Y, Zhang B, Liu S (2013) Research on urban water demand forecasting method based on fuzzy neural network based on T-S model. J Saf Environ (2)
9. Sang YJ et al (2010) T-S fuzzy neural network algorithm application in nonlinear contral. In: 2010 ISECS international colloquium on computer, communication, control, and management (CCCM).IEEE, pp 24–27

Development on Management System of Automated High-Rise Warehouse for Mid-Small Enterprises Based on Django

Chuanhong Zhou, Chao Dai, Pujia Shuai and Fei Qi

Abstract This article mainly presents a management system of automated high-rise warehouse and one specific design of the communication mechanism. The B/S architecture system based on Django can complete in/out stock, inventory taking, inventory transfer operations etc. In addition, the automation of warehouse management can be directly realized through the information interaction in order to improve the accuracy and convenience of storage management in enterprises.

Keywords Automated high-rise warehouse · Django · Communication

1 Introduction

With the rapid economic development, it's necessary for Chinese economic development to establish the developed modern logistics. Furthermore, Storage system has become indispensable component of freight hub and distribution center in the logistics system. So it is very important to develop and establish automated high-rise warehouse. With the continuous expansion of the industry market, production scale, sales market and the rapidly increase of inventory in small or medium enterprises, the original common warehouse has been unable to meet the needs of the company's development. So it often appears that the backlog of finished, expired or damaged goods, as well as the slow pace in/out of stock, inaccurate goods in/out stock, and even it seriously affects the economic benefits of enterprises and customer service quality [1].

Currently, the significance of setting up the automatic high-rise warehouse which can be applied to small and medium enterprises is distributed mainly in the following aspects: warehouse storage is customer-centered and requiring for a variety of stored goods, automation of storage devices, the rapidity of warehousing

C. Zhou · C. Dai (✉) · P. Shuai · F. Qi
School of Mechatronic Engineering and Automation, Shanghai Key Laboratory of Intelligent Manufacturing and Robotics, Shanghai University, 200072 Shanghai, China
e-mail: uuihop@163.com

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_7

operation, the operation of electronic and miniaturization of orders. Meanwhile, with the development of modern industrial production, Flexible Manufacture System (FMS), Computer Integrated Manufacture System (CIMS) and factory automation (FA) put forward higher requirements for automated warehouse so that handling storage technology will have more reliable, more real-time information, factories and warehouses in logistics must be accompanied by parallel flow of information.

2 Introduction and Design of Management System of Automated High-Rise Warehouse

Before the practical design system, we need to conduct a requirement analysis on automated high-rise warehouse management software as well as combining with the demand analysis and system architecture. Through the reasonable analysis, we can make the system do what we want.

2.1 Requirement Analysis

Before developing system, it's necessary to have a detailed analysis on the functional requirements of the system and specify the problem domain, this is the basic work of software system development. The management system is mainly composed of two parts and they are the automatic recognition and acquisition system. This paper mainly researches on the information management of automated high-rise warehouse. However, its implementation and scheduling mechanism of the warehouse system are omitted. The requirements of this system include the following several aspects:

- A. Through the scientific and reasonable storage management of products to ensure that the production operation and achieving the goal of improving the ability of responsiveness and efficiency.
- B. Including functions of general information management system of rights management, data query, statistics management etc.
- C. Standardizing business processes, the adoption of scientific management processes and organizational structure to make inventory management optimization, no redundancy, basic attribute of parallel operation to ensure the effective implementation of business processes.
- D. Implementing automation of completing in/out stock & inventory taking operations and reducing the cost of inventory management. The inventory information could be automatically collected, implementation of inventory management of paperless operation can directly improve the efficiency of process of in/out stock so as to increase the inventory of the center for throughput

which can both improve the quality of the inventory operation and improve the accuracy of the goods query the goods to make the system automatically prompt warning and provide timely inventory information to managers and decision makers when the stock inventory level below the safety standard.

- E. Reducing the workload of the warehouse personnel and improving the work efficiency, meeting the need of the development of modern inventory.

2.2 System Analysis [2]

This system is divided into three layers which are management layer, monitoring layer and execution layer. Management layer is the center of the warehouse management system, its main function is responsible for the automated multi-layered storehouse system information management and coordination of the normal operation of the system equipment. Monitoring layer is an important part of automatic high-rise warehouse system and it's also an information transfer part of the system which is mainly used to receive the instructions from management and control the logistics equipment to complete the required tasks according to the instructions. Execution layer consists of some automated logistics machines.

2.3 Functional Selection

Based on the three-tier architecture, the warehouse management system will use two separate services (ERP integration service, PLC integration service) to complete the transmission of enterprise-class ERP product information docking and the implementation of the main system of the main PLC command transmission. The architecture is shown in Fig. 1.

1. Using a separate ERP integration services. This design makes different ERP systems only need to set aside the ERP system interface and ERP integration implement system functions where there is a simple modification of it which can avoid Re-development of the warehouse management system because of the ERP system replacement in the ERP system for the replacement of the warehouse management System development. Besides, it can also ensure that the system have high portability between different enterprises.
2. Using a separate PLC integrated service. The system architecture can complete these tasks through a separate PLC integrated service and the system only need to be stored and read the database information so it can send commands and receive feedback so that when the PLC technology development, we only need to modify the PLC integration services (Such as reading the database from a single read to multi-task read ways, etc.) can complete the system upgrade, but

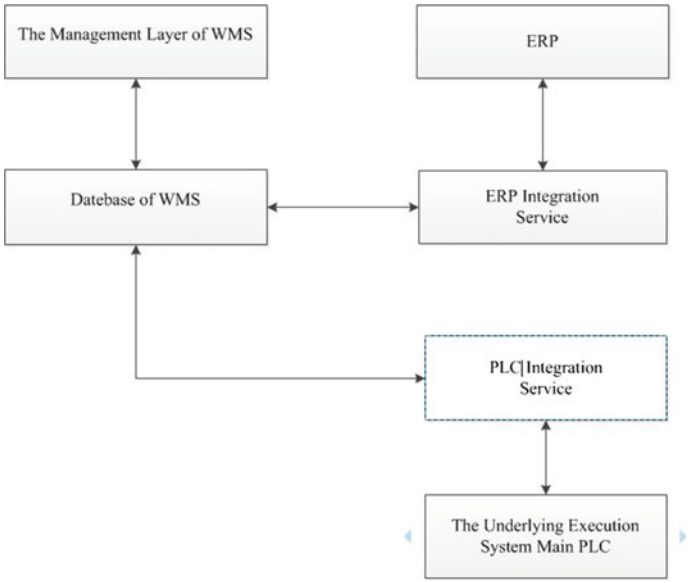


Fig. 1 The system integration services architecture

do not need to make changes to the entire system, which will save the cost of the enterprise.

3 Modularization Development of Management System of Automated High-Rise Warehouse

According to the last section about requirement analysis of the warehouse management system and combining with system architecture, we can conclude the functional analysis of warehouse management system and think up the development modules of the whole warehouse management system. This section is mainly discussed about module development.

3.1 Introduction of Development Framework

The system is developed with Django framework [3] and it is an open source Web application framework based on Python language. The basic conception of the framework is the rapid development and a concise and practical design. Furthermore, the framework itself brings great convenience to the program development so that programmers could be more focused on the Web application itself.

3.2 Part of the Program Design of Management System of Automated High-Rise Warehouse

The design of the warehouse management system including in stock module, out stock module, stock transfer, inventory taking module, log module and so on. In all of these modules, the log module is the most important part. The storage interface development processes combined with different situations in the in/out stock choice strategy are mainly discussed in this paper.

The main design of the system has three ways of in stock operations, namely, one-way in stock, round in stock, empty pallet in stock and the main way to achieve is as the following:

(1) one-way in stock: specifying the state of storage space which is empty or existing goods and executing this command can directly sent goods waiting in the entrance to the target location.
(2) round in stock: from the storage: specifying the state of storage space which exists a tray on it and once executing this command the stacker will firstly remove the tray from the target location and then the items are placed in the trays in the storage operation.
(3) empty tray in stock: specifying the state of storage space which is empty or existing goods and when executing this command, the stacker storage will sent the empty tray waiting in entrance to the designated target location.

For in stock operations, firstly, we should select goods that are waiting for entering in the warehouse on the operating interface. The in stock operation can not do without goods in parts list. If the target goods are found in parts list, the project name, product barcode, warehouse voucher and numbers of goods will be filled in. After confirming the products number and the states in various storage spaces. If there exists appropriate tray with some goods on it then we can select the location for the target location and select “round in stock” task type.

The specific in-stock operation algorithms is shown as following and the system provides “one-way in-stock”, “round in stock”, “empty pallet in-stock” these three in stock mode. The specific mode determined by the warehouse managers will depend on the size, the number, the volume and shape by the artificial operation. After determining the workflow and the information which are required for in-stock module, we need to collect and sort the data and put them in database. The storage process requires module and information. Next, the overall structure of the code in the form of a model-template-view will be illustrated in the following figure. Firstly, we should design the database forms and declare their fields and attributes through the ORM database in mode.py package [4]. The following step is the design of web page including the layout of the interface design and the determination of interface content display. In this procedure, we adopt HTML language to design the interface how to display and Bootstrap framework which is used to design the layout [5]. Otherwise, for specifying display content, it depends on the view file. The view file will be filtered through database data and the records that

meet the constraints will be taken out from the filtering and then these data will be stored in the form of a dictionary and the system will send a request to the URL address [6]. At the same time, the form file through the type definition will read the data from the database and offer the information to the template file [7].

In fact, the warehouse information processing of the automate warehouse system are the addition, deletion, update, selection operations of database information and this system has adopted the PostgreSQL database [8].

3.3 *Communication Protocol Between PLC and the Monitor [8]*

On the basis of TCP/IP protocol, the establishment of a set of methods of communication will bring great convenience to the communications between warehouse management system (WMS) [9] and warehouse execution system (WES) in order to play a major role of controlling the lower computer with the upper computer [10].

The TCP/IP protocol instruction frame is named after WDTcpChatter command frame as shown in Table 1 and its structure is as follows: WDTcpChatter instruction frame is used for the unified data communication between the client and the server. Description of command in detail is shown Table 2. Its basic structure is [11].

Table 1 WDTcpChatter command frame

WDTcpChatter command field								Data field
Head	Mode	SendID	RecvID	Resv1	Resv2	Resv3	Resv4	Data field
1 byte	1 byte	1 byte	1 byte	1 byte	1 byte	1 byte	1 byte	2 or 6 bytes

Table 2 Details of WDTcpChatter command

WDTcpChatter command frame	
Head	0 × 01, //valid command starting flag (command header)
Mode	0 × 01 as polling, 0 × 02 as response
SendID	0 × 01 as main PLCID
RecvID	0 × FF as monitoring end ID
Resv1 ~ Resv4	All reserved for 0 × 00
WDTcpChatter data field	
Data 1	1 byte (as respond sign, 0 × 08 as normal end command response, 0 × 00 as non response, 0 × FF as errors)
Data 2	1 byte (as error indication for the underlying control system, 0 × 00 as normal)

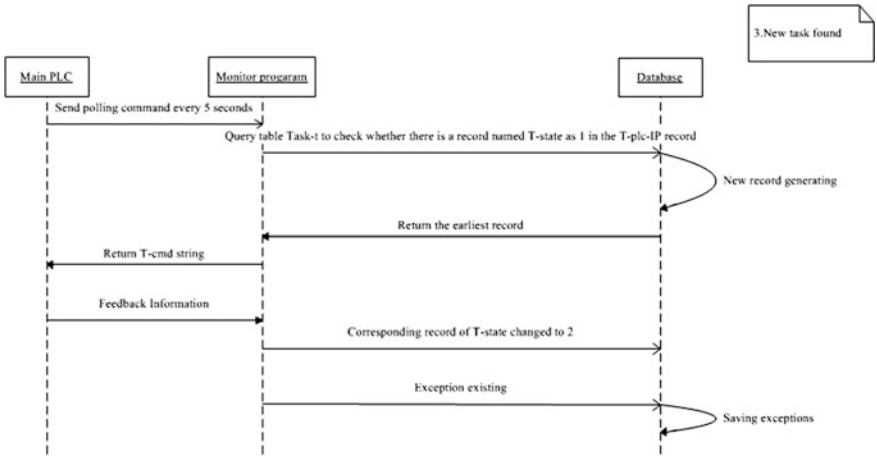


Fig. 2 Interaction between the monitor and PLC

WDTcpChatter command + WDTcpChatter data fields, including:

The composition of WDTcpChatter command field: command head + command request mode, the sender ID + ID + reserved command. The length of WDTcpChatter is 8 bytes. The length of WDTcpChatter data field: (A) The length of command that the main PLC sending to monitoring terminal is 2 bytes; (B) The length of command that the monitoring end sent to the main PLC is 6 bytes. Besides, the interaction between the monitor and PLC as shown in Fig. 2.

3.4 Introduction of in-Stock Module in Management System of Automated High-Rise Warehouse

This section will take the storage module as an example and the in-stock operation process of the automated high-rise warehouse as well as the realization of the function and interface display were introduced in this paper.

The in-stock task mainly involves the following operation flow: selecting in-stock type of parts—generating a warehousing entry—generating in-stock task—sending in-stock command to the PLC end—receiving the PLC response. The operation of the UML sequence diagram [12] as shown in Fig. 3.

The main interface of the in-stock operations are as follows in Fig. 4.

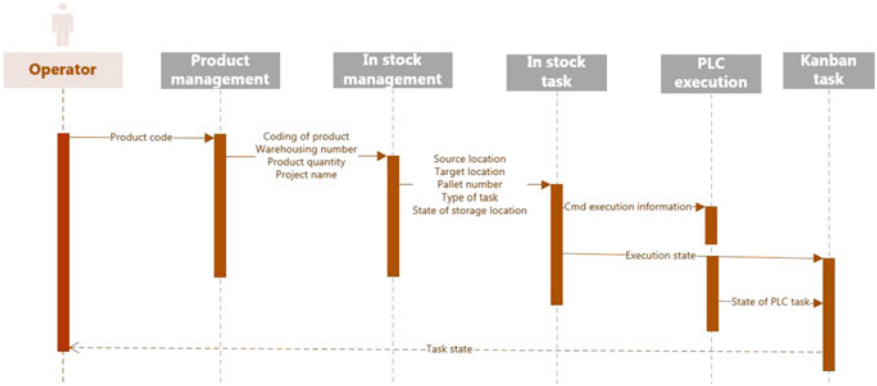


Fig. 3 UML sequence diagram

Fig. 4 Kanban of completing the in-stock task

4 Conclusions

This paper mainly introduces the development of a management system server of automated high-rise warehouse as well as the design of the communication system. The design of WMS (Warehouse Management System) software based on Django can easily handle a series of operational requirements in Mid-Small enterprises and the development of communication system based on the Twisted framework can implement automatic operation between the management system and the lower computer so that the accuracy of and rapidity of warehousing logistics are greatly improved.

Acknowledgements This work is supported by subject of Science and Technology Commission of Shanghai Municipality in key technology development and demonstration application of the unmanned factory for industrial robot production (No. 14DZ1100700). We thank Shanghai Key Laboratory of Intelligent Manufacturing and Robotics for assistance with the automotive parts cost data.

References

1. First Annual Intuit QuickBase EMPOWER (2015) Conference spotlights citizen developers using technology to solve their own business challenges "New York Times" Intuit QuickBase Django FAQ. Retrieved from 2 Sep 2014
2. Nicole R (2016) FAQ: General—Django documentation—Django. Title of paper with only first word capitalized. J Name Stand Abbrev. Retrieved 30 Apr 2016, in press
3. Bootstrap 3.3.7 Released (2016) Retrieved 30 July 2016
4. Otto M (2015) Bootstrap 4 alpha. Bootstrap blog. Retrieved 20 Aug 2015
5. Pavel S (2016) Common problems that you can solve with a Bespoke Warehouse App. Magora Sys. Retrieved 18 May 2016
6. Guo F, Liu WL, Cao Y, Liu FT, Li ML (2016) Optimization design of a webpage based on Kansei Engineering. Hum Factors Man 26(1)
7. Workflow management coalition. Workflow management coalition home page
8. Ehrig H (2003) Petri net technology for communication-based systems: advances in Petri Nets. Springer, p 323. ISBN 978-3-540-20538-8
9. McCrea Bridget (2002) WMS update: what do we need to run a warehouse? Logistics Manage 2016:556
10. An innovative workflow mapping mechanism for grids in the frame of quality of service (2008). Future Gener Comput Syst 24(6):498–511 (Elsevier.com)
11. Li J, Peng Z, Gao S, Xiao B, Chan H (2016) Smartphone-assisted energy efficient data communication for wearable devices. Comput Commun
12. Bashir RS, Peck Lee SP, Khan SUR, Chang V, Farid S (2016) Uml models consistency management: guidelines for software quality manager. Int J Info Manage

Underwater Image Enhancement by Fusion

Can Zhang, Xu Zhang and Dawei Tu

Abstract Underwater images are suffered by several factors like color absorption, light scattering, and these lead to poor visibility and low contrast of an underwater image. Typical image restoration methods rely on the popular Dark Channel Prior. However, the result of the underwater image which has only been dealt with the restoration algorithm is still not ideal. In order to get better results, we introduce a new underwater image enhancement approach based on multi-scale fusion strategy in this paper. In our method, we first obtain the restored image on the base of underwater image model. Then we get the white balance and contrast enhancement image of the restored image respectively. Finally, these two derived inputs are blended by multi-scale fusion approach, using saturation and contrast metrics to weight each input. This algorithm reduces the execution time and can effectively enhance the underwater image. The experimental results demonstrate that our method can obtain better visual quality.

Keywords Image restoration · Image enhancement · Image fusion

1 Introduction

The ocean is an important part in the earth, thus countries in the world are all committed to the study of the marine environment. In order to make full use of the marine resources, we must have a deep understanding on the ocean, so it is very important to get all kinds of information of the ocean. As the carrier of information, image is one of the important means to obtain information of underwater world. At present, underwater imaging technology has been applied in many fields, such as marine biology research, inspection of the underwater infrastructure and the control of underwater vehicles. However, underwater scenes are usually veiled by the light

C. Zhang (✉) · X. Zhang · D. Tu
Department of Mechanical Engineering and Automation,
Shanghai University, Shanghai, China
e-mail: 475537062@qq.com

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_8

interaction with the medium: absorption and scattering of the light induce poor contrast, low luminosity and restricted visibility. Thus the images acquired from the water often suffer from visual degradation and have to be preprocessed before analysis or using.

How to get high quality underwater image is a project that has been studied by scholars all over the world. Underwater image enhancement methods can be divided into three categories, the first category is based on underwater image degradation model image, the second method is based on the contrast and histogram equalization method and the third is based on image fusion method.

The first method uses image degradation model [1] to enhance the underwater image, by studying on physical process of natural light in the water medium to establish the degradation model of underwater image, then inverse the process of degradation image to obtain the final image. In the process of parameter estimation, some studies use the calibration plate to get the model parameters directly. For example, Roser et al. [2] use sparse 3D mapping of original degraded image to calculate visual properties. By the influence of He et al. [3] on the image defogging method, many researchers use the dark channel prior method to deal with the underwater images. Yang et al. [4] use image enhancement method based on dark channel prior with low complexity. In his paper, the median filter is used to estimate the transform function of the input image, and the light of the atmosphere is calculated by the dark channel. Li et al. [5] use a new fast channel prior method to deal with the underwater image or video, which can effectively reduce the image noise and improve the brightness of the dark area.

The second method we mentioned above is based on the contrast and histogram equalization method. Contrast extension extends the contrast of the image to the lower and upper thresholds and histogram equalization map gray value to a unified system to adjust the bright or dark areas. There are many improved methods based on histogram equalization. For example, Hitam et al. [6] put forward an adaptive histogram equalization with mixed constraint contrast, dealing in the RGB color space and HSV color space respectively, and finally mix using Euler norm. Karam et al. [7] use the classical contrast enhancement technique and the fuzzy histogram equalization method to study the image with different depths, this method converts the RGB image to a Lab image and extends the original value in the Lab space.

There are some methods combining image restoration with image enhancement together. In order to achieve high quality underwater image, the combination of various image enhancement techniques is widely used. Compared with the single image enhancement method, the results show that the method has better performance in dealing with the main problems such as noise, low contrast, uneven illumination, color fading and so on. Ancuti et al. [8] put forward a fusion technique for underwater image and video. In his paper, he first process the underwater image with white balance method, and then use bilateral filtering and other methods to remove the noise and correct the color, finally fuse them by using image pyramid.

In this paper we propose a fusion-based strategy [9] that enhance underwater image with high efficiency and low complexity. Following the optical underwater model [1], we first obtain the restored image. Then, since the result of the

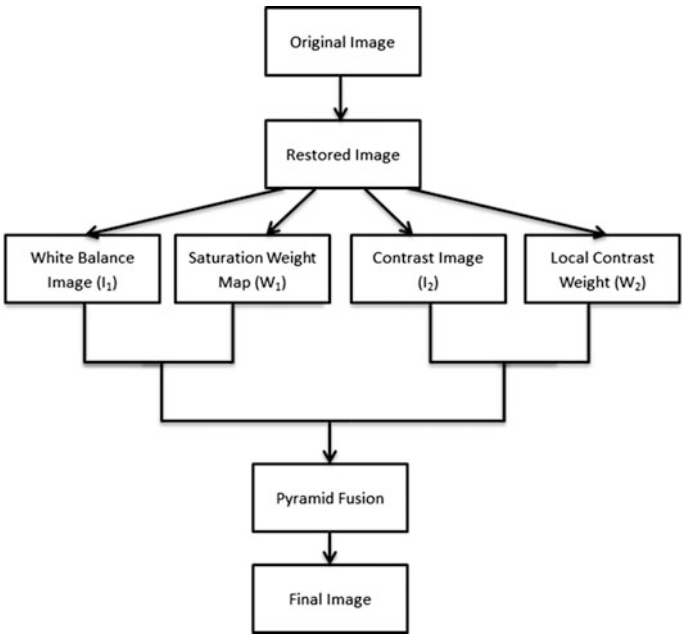


Fig. 1 The outline of our algorithm

underwater image which has only been dealt with the restoration algorithm is still not ideal, we compute the white balance image and the contrast enhancement image of the restored image. Last, these two images derived from the restored image, guided by several quality measures, are seamlessly blended by using a multi-scale fusion approach. The outline of our algorithm is showing in Fig. 1.

2 Underwater Image Model

According to the common and popular optical model of McGlamery, the total radiance of an image that reaches observers can be modeled as two components: the direct component E_D and the back-scattering component E_{BS}.

The direct component E_D represents the direct reflection of light from the object, and is expressed at each image coordinate (x, y) as:

$$ED(x, y) = J(x, y)\mathrm{e}^{-\eta d(x,\, y)} = J(x, y)\; t(x, y) \tag{1}$$

where $J(x, y)$ is the radiance of the object, $d(x, y)$ is the distance between the observer and the object, and η is the attenuation coefficient. The exponential term $\mathrm{e}^{-\eta d(x,\, y)}$ is also known as the transmission $t(x, y)$ through the underwater medium.

The back-scattering component E_{BS} is known as the reflection from the particles of the medium, this component can be expressed as [10]:

$$EBS(x, y) = B_{\infty}(1 - \mathrm{e}^{-\eta d(x, y)}) \tag{2}$$

where B_{∞} is a scalar known as the back-scattered light or the water background [11].

Incorporating these two components and we can get the simplified underwater optical model as follow:

$$\begin{aligned} I(x, y) &= J(x, y)\mathrm{e}^{-\eta d(x, y)} + B_{\infty}(1 - \mathrm{e}^{-\eta d(x, y)}) \\ &= J(x, y)t(x, y) + B_{\infty}(1 - t(x, y)) \end{aligned} \tag{3}$$

Transform the inverse of the Eq. (3), and then we can get the restored image $J(x, y)$, as shown in the Eq. (4):

$$J(x, y) = \frac{I(x, y) - B_{\infty}}{t(x, y)} + B_{\infty} \tag{4}$$

There are two unknowns in the image acquisition model defined by Eq. (4): the transmission map $t(x, y)$ and the back-scattered light B_{∞}.

2.1 Transmission Estimation

Following the paper of He et al. [12], we approximate the transmission t(x, y) based on the Dark Channel Prior. This prior assumes that natural objects have a weak reflectance in one of the color channels. That means the direct radiance is small, or dark, in at least one of the RGB color channels. Given this assumption, the transmission map t(x, y) can be estimated from the weakest color over a neighborhood of (x, y). So the equation can be shown as follow:

$$t(x, y) = 1 - \min_{pix \in \Omega(x, y)} \left(\min_{c \in \{r, g, b\}} (I^{c}(x, y)) \right) \tag{5}$$

where $\Omega(x, y)$ represents a local patch centered at (x, y).

2.2 Back-Scattering Estimation

According to Ancuti's paper [13], he introduces an alternative approach to be able to adapt locally to the back-scattered light. Their proposal achieves visually

improved de-scattering, so we use this method to estimate back-scattering in our paper, the equation shown as follow:

$$B_{\infty}^{c}(x,y) = \max_{(x_*,y_*)\in M_{DC}^{\psi(x,y)}} \left(\min_{pix\in\Omega(x_*,y_*)} \left(I^{c}(pix_x, pix_y) \right) \right) \tag{6}$$

where Ψ(x) is a square patch centered at x. As for the global estimator, $M_{DC}^{\Psi(x)}$ denotes the set of positions in Ψ(x) where the dark channel is max, i.e. reaches the value of its local maximum over Ψ(x). In practice, the patch Ψ is typically larger than Ω. We use a default value of 2 for the ratio between the sizes of the Ψ and Ω patches.

2.3 Image Restoration Method

It is now known that the original image $I(x,y)$, the transmission map $t(x,y)$ and the back-scattered light B_{∞}, put these three known quantities into the formula (4), and we can obtain a clear recovery image.

3 Image Enhancement by Multi-scale Fusion

We have already obtained the restored image by using underwater image restoration algorithm which has been mentioned in the previous section. But the problems of the color distortion and low contrast of the image are not solved, so in order to achieve high quality enhancement images, we propose a fusion-based strategy. By choosing appropriate weight maps and inputs, our fusion-based method is able to effectively enhance underwater images.

3.1 Inputs

(1) White balance

Due to the different wavelengths of light, light absorption and scattering would lead to some of the shorter wavelength light scattering, and the rest light of other length wavelength will cross the medium. This will inevitably lead to the deviation of the color. Similarly, the absorption of light will weaken the intensity of light, so that the formation of the image color of light energy is reduced, so there is a color deviation between underwater image and the real scene, in order to restore the color of the image, it is very important to use a simple and efficient method to deal with the white balance.

We use the dynamic threshold white balance method to deal with three groups of underwater images, the results are shown in Fig. 2.

Fig. 2 **a** Restored image. **b** White balance of the restored image

(2) Global Contrast

In addition to the color distortion caused by the attenuation of light, the backscattering from the suspended particles during the propagation of the optical image reduces the brightness of the image. In order to get a clear image, it is necessary to improve the global contrast of the source image.

In this paper, we use the histogram equalization to deal with three groups of underwater image, the results are shown in Fig. 3.

Fig. 3 **a** Restored image. **b** Contrast enhancement of the restored image

3.2 Weight Maps

Inspired by fusion underwater approach [14], we derive two weight maps. This ensures that locations of high saturation or high contrast will receive greater emphasis in the fusion process.

(1) Saturation weight map

Saturation is also known as the purity of the color, the greater the saturation, the more vivid the image color, people usually want the image contains a high degree

of saturation, so that the visual effect is better. Hence, we use saturation weight for the final fusion. Saturation weight method is as follows:

$$W_1(x,y) = \text{std}(I^r(x,y)I^g(x,y)I^b(x,y)) \tag{7}$$

As we can see from the above, the saturation weight is the standard deviation of the value of each pixel in the three channels RGB of the input image.

(2) Local contrast weight

Contrast weight is used to detect the local change of the input image, and can get better edge information and texture changes. The contrast weighted formula is as follows:

$$W_2(x,y) = |Laplacian(I^r(x,y))| + |Laplacian(I^g(x,y))| + \left|Laplacian(I^b(x,y))\right| \tag{8}$$

We can first obtain the Laplacian result of the three channel of the original image, and then calculate the L1 norm for each pixel (the sum of the absolute values of each element).

3.3 Multi-scale Fusion

The reflection of multiresolution of an image can be achieved by using image pyramid. Image pyramid is an effective but simple image structure. The principle of image pyramid method is that: each of the participating fusion images is decomposed into multi-scale pyramid image sequences, and the low resolution images in the upper layer, the high resolution image in the lower layer, the upper image size is 1/4 of the previous image. Put all the images of pyramid in the corresponding layer into blending with a certain rule, and we will get fusion pyramid.

In the above, we have got two input images and their corresponding weights, the two input images will be multiplied by their corresponding weights, and then blended by multi-scale fusion, so that we can get the final results.

In order to facilitate the subsequent weighted fusion, each weight map must be normalized first, and the equation can be shown as follow:

$$\begin{aligned} \overline{W}_1(x,y) &= \frac{W_1(x,y)}{W_1(x,y) + W_2(x,y)} \\ \overline{W}_2(x,y) &= \frac{W_2(x,y)}{W_1(x,y) + W_2(x,y)} \end{aligned} \tag{9}$$

The two input images (white balance image I_1 and contrast image I_2) multiplied by the corresponding weights respectively, the output image can be obtained:

$$F(x,y) = \overline{W}_1(x,y)I_1(x,y) + \overline{W}_2(x,y)I_2(x,y) \tag{10}$$

However, in order to further improve the quality of the output image, we don't use linear fusion method in this paper, we get the final image based on the pyramid fusion method. In this paper, the input image I is processed by Laplacian Pyramid, and the normalized weight W is processed by Gaussian Pyramid, and the fusion is carried out hierarchically. Since Laplacian Pyramid and Gaussian Pyramid have the same number of layers, the fusion of the input image and the corresponding weights is also carried out in each layer. Specific formula is as follows:

$$\begin{aligned} F^l(x,y) &= \sum_{k=1}^{k} \left\{ G^l(\overline{W}_k(x,y)) L^l(I_k(x,y)) \right\} \\ &= G^l(\overline{W}_1(x,y)) L^l(I_1(x,y)) + G^l(\overline{W}_2(x,y)) L^l(I_2(x,y)) \end{aligned} \tag{11}$$

where l represents the number of the pyramid levels and $L(I)$ is the Laplacian version of the input I while $G(W)$ represents the Gaussian version of the normalized weight map of the W. This step is performed successively for each pyramid layer, in a bottom-up manner. The final output image F is obtained by summing the fussed contribution of all inputs.

The final enhanced underwater image is shown in Fig. 4.

4 Experimental Result and Discussion

In the section three we use dynamic threshold white balance method to obtain the white balance input I_1 and use the histogram equalization method to achieve contrast input I_2. In Figs. 5 and 6, we compare our methods with several common algorithms.

From Fig. 5, we can see that the dynamic threshold white balance method (the white balance method used in this paper) is more suitable for human visual habits. Gray world method has color deviation, and the brightness of the result of the perfect reflector method is not improved. Therefore, the white balance used in this paper is more suitable for dealing with the impact of color fading in underwater image.

Figure 6 presents the comparison of several contrast enhancement methods. We can see that the histogram equalization algorithm is the best. The results based on the Laplace operator and Log transform are too exposed, they may be more suitable for image enhancement with low contrast and low gray value; The results obtained by image enhancement based on gamma transform are dark, so it also not suitable for our underwater image.

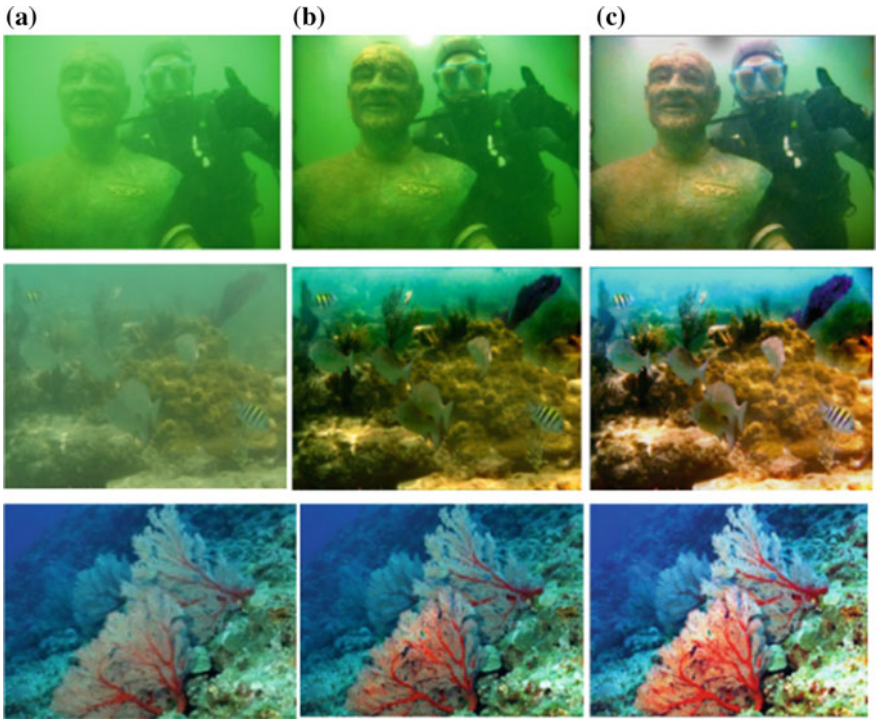


Fig. 4 **a** Original image. **b** Restored image. **c** Enhanced image

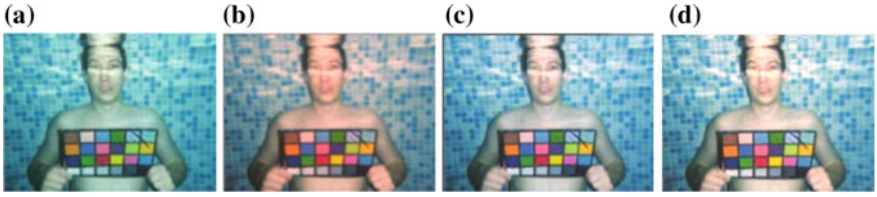


Fig. 5 The comparison of white balance methods. **a** Original image. **b** World gray method. **c** Perfect reflector method. **d** Dynamic threshold (The method we used)

Furthermore, in order to illustrate the performance of our method, we use two groups of test images, shown in Figs. 7 and 8. Comparing these images by two enhancement methods, we can obviously see that our method can effectively enhance the underwater image.

The literature [15] employs underwater image enhancement based on dark channel prior. The method firstly use the median filter to estimate depth map, then combine the black channel to build scene image, finally, they employ the color enhancement. However, this method exist estimation error. And it will result in the deviation of the color fidelity. The comparison result is shown in Fig. 7.

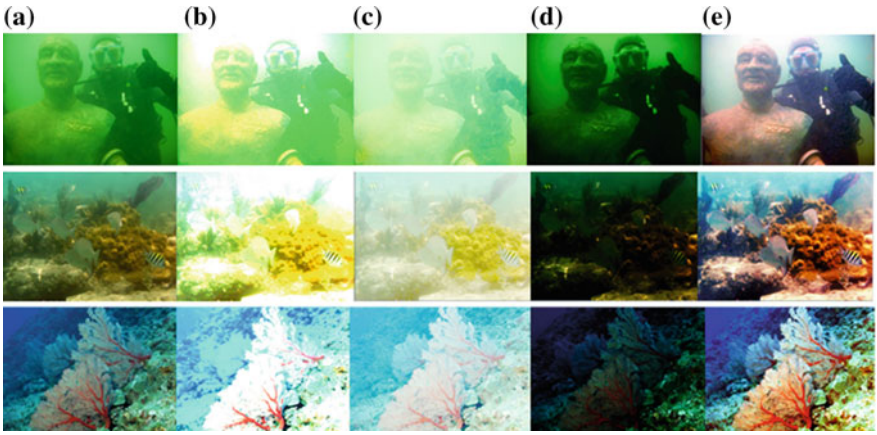


Fig. 6 The comparison of contrast enhancement methods. **a** Original image. **b** Laplace method. **c** Log method. **d** Gama method. **e** Histogram equalization method (The method we used)

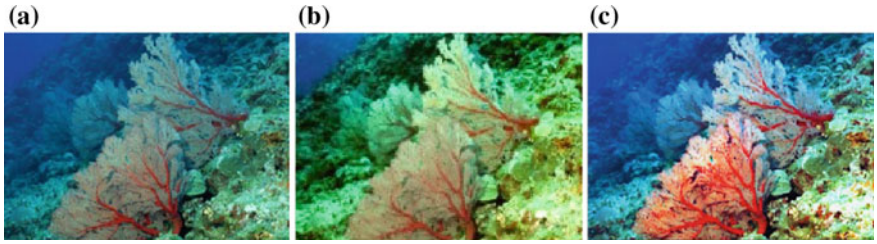


Fig. 7 **a** Original image. **b** literature [15]. **c** Our result

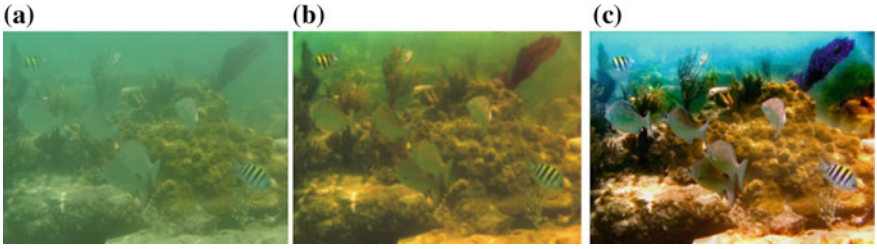


Fig. 8 **a** Original image. **b** literature [16]. **c** Our result

Seen the group of comparison in Fig. 8, whether the degree of bright color or the display of details, our result is more satisfactory.

5 Conclusion

In this paper, we introduce a new underwater image enhancement approach based on multi-scale fusion strategy. From the experiment, we can see that our method shows better results compared to other enhancement methods, no matter in colour or contrast. This algorithm reduces the execution time and can effectively enhance the underwater image.

Acknowledgements This research was partially supported by the National Nature Science Foundation of China (Grant no. 51575332 and no. 61673252) and the key research project of Ministry of science and technology (Grant no. 2016YFC0302401).

References

1. McGlamery BL (1980) A computer model for underwater camera systems. Ocean optics VI. Int Soc Optics Photonics 221–231
2. Roser M, Dunbabin M, Geiger A (2014) Simultaneous underwater visibility assessment, enhancement and improved stereo. In: IEEE international conference on robotics & automation, pp 3840–3847
3. He K, Sun J, Tang X (2010) Single image haze removal using dark channel prior. In: 2013 IEEE conference on computer vision and pattern recognition. IEEE, pp 2341–2353
4. Yang HY, Chen PY, Huang CC et a1 (2011) Low complexity underwater image enhancement based on dark channel prior. In: 2011 Second international conference on innovations in bio-inspired computing and applications (IBICA). IEEE, pp 17–20
5. Li Y, Lu H et a1 (2015) Underwater image enhancement using inherent optical properties. In: 2015 IEEE International conference on information and automation. IEEE, pp 419–422
6. Hitam MS, Yussof WNJHW, Awalludin EA et al (2013) Mixture contrast limited adaptive histogram equalization for underwater image enhancement. In: 2013 International conference on computer applications technology (ICCAT), pp 1–5
7. Karam GS, Abood ZM, Saleh RN (2013) Enhancement of underwater image using fuzzy histogram equalization. www.Research.ijais.org
8. Ancuti CO, Ancuti C, Haber et a1 (2011) Fusion-based restoration of the underwater images. IEEE International conference on image processing. IEEE, pp 1557–1560
9. Ancuti CO, Ancuti C, Bekaert P (2010) Effective single image dehazing by fusion. IEEE
10. Schechner Y, Karpel N (2005) Recovery of underwater visibility and structure by polarization analysis. IEEE J. Oceanic Eng
11. Lythgoe JN, Hemmings CC (1967) Polarized light and underwater vision. Nature 213(79)
12. He K, Sun J, Tang X (2009) Single image haze removal using dark channel prior. IEEE CVPR
13. Ancuti C, Ancuti CO (2016) Multi-scale underwater descattering. ICPR, pp 4202–4207
14. Ancuti C, Ancuti CO, Haber T, Bekaert P (2012) Enhancing underwater images and videos by fusion. In: IEEE conference on computer vision and pattern recognition (CVPR)
15. Yang HY, Chen PY, Huang CC et al (2011) Underwater image enhancement based on dark channel prior. IEEE computer society, pp 17–21
16. Drews P, Nascimento E, Moraes F, Botelho S, Campos M, Grande-Brazil R (2013) Transmission estimation in underwater single images. IEEE Workshop ICCV

Restoration and Enhancement of Underwater Light Field Image

Wei Cui, Chen Li, Can Zhang and Xu Zhang

Abstract In this paper, we propose an algorithm for underwater light field image restoration and enhancement. The algorithm is based on the image haze removal using dark channel prior and Pyramid image fusion. In the process of restoration and enhancement, the light field image is converted into four-dimensional data, and each angle image of the four-dimensional data is restored and enhanced. After this, the four-dimensional data is refocused and all-focused again to get the clearer image. Final experimental result shows that the proposed algorithm has good effects on the restoration and enhancement of underwater light field image.

Keywords Dark channel prior · Fusion · Refocus · All-focus · Four-dimensional data

1 Introduction

Ocean is an important part of earth and there are lots of energy and resources. Many countries in the world are committed to the study of the marine environment. In order to make a better use of the marine resources, we must cultivate a deep understanding of the ocean and get the information as much as possible. However, in the complex environment of the ocean, underwater imaging is seriously affected due to the absorption of the incident light by pure water, inorganic salts and a large amount of soluble organic matter, and the scattering of light by suspended components such as zooplankton, plants, bacteria, detritus and mineral particles. Therefore, it is very necessary for us to do some restoration and enhancement on underwater image so that we can get more visual information.

W. Cui (✉) · C. Li · C. Zhang · X. Zhang
Department of Mechanical Engineering and Automation,
Shanghai University, Shanghai, China
e-mail: 18301926762@163.com

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_9

In recent years, a lot of researches have been conducted on the underwater image preprocessing methods. The computer model of the classical underwater imaging system was proposed by McGlamery [1] in 1979. On the basis of the classical underwater imaging system, Andreas et al. [2, 3] has put forward the framework of underwater image preprocessing, they used the convolution method of Gauss function to estimate the backscattered light component. However, one of the problems of this algorithm is that the estimation of scattered light is much too time-consuming. Liu [4] used the dark channel theory to estimate the depth of field information of underwater environment, and then did the recovery processing according to the underwater imaging model. With the development of light field technology, light field camera is more and more popular in our life and many people use it to catch images. Light field image is the fusion of many kinds of images, and it is also the hybrid alignment of orientation image and position image [5]. So when we get the light field image that we actually get an angle image of it. The light field image not only records the brightness of the spatial position, but also the direction of the light [6, 7] which means that the image can be refocused [8, 9] according to the information of the light field. With the great advantages, light field camera is more and more widely used in underwater imaging. However, there are still some problems in underwater imaging for light field camera. Due to some factors, underwater light field image is fuzzy and distorted that the image needs to be restored and enhanced to get more information of the ocean.

In this paper, underwater light field image is restored and enhanced on the basis of the dark channel theory [10] and Pyramid image fusion [11, 12]. With the properties of light field image, the light field image can be converted into four-dimensional data [6, 13] which records the orientation and position of the image. When getting the orientation and position, it is able for each angle image of the four-dimensional to be restored and enhanced. Besides, these different angle images also have the property to be merged into four-dimensional data again which can be refocused and all-focused [14–19] to get the final image. In the experiment, the light filed image from Lytro is restored and enhanced. Final result shows that the proposed algorithm has good effects on the fuzzy underwater light field image.

2 Restoration and Enhancement of Light Field Image

When restoring and enhancing the underwater light field image, it is a must to build the underwater light field image model first. Based on the dark channel theory and Pyramid image fusion, a new model is proposed. And with the new model, final clearer image can be obtained by the further processing of refocus and all-focus.

2.1 Underwater Light Field Image Model

Underwater image model is based on a computer model of underwater imaging system proposed by McGlamery, the underwater image formation can be simulated from the model. He believes that in underwater media, the viewer can see the image of the total emissivity is composed of three components: directly attenuating rays component E_D, forward scattering component E_{FS} and backscattering component E_{BS}. These components are mainly caused by the radiation of the objects in the scene and the ambient light.

The energy of directly attenuating ray component E_D index decays with depth, the expression is as follow:

$$E_D(x,y) = J(x,y)\ \mathrm{e}^{-\eta d\,(x,y)} = J(x,y)t(x,y) \tag{1}$$

where $J(x,y)$ is the radiant brightness of an object, $d(x,y)$ is the distance between the observer and the object, η is the attenuation coefficient, $\mathrm{e}^{-\eta d\,(x,y)}$ is known as the propagation of an underwater medium and we can express it as $t(x,y)$.

The forward scattering component E_{FS} is a part of the incident light deflection. Generally speaking, it has little effect on the whole image degradation process, so it can be ignored.

The backscattering component is the main reason for the loss of contrast and color deviation. Within a reasonable distance (usually 3–10 m), the component can be expressed as follow:

$$E_{BS}(x,y) = B_\infty(1 - e^{-\eta d\,(x,y)}) \tag{2}$$

where B_∞ is called backscattered light or background light. In the case of uniform illumination, the illumination can be considered as an equivalent light source from infinity.

Combining these components and ignoring the forward scattering, the underwater imaging model can be obtained as follow:

$$\begin{aligned} I(x,y) =& J(x,y)\mathrm{e}^{-\eta d\,(x,\,y)} + B_\infty(1 - \mathrm{e}^{-\eta d\,(x,\,y)}) \\ =& J(x,y)t(x,y) + B_\infty(1 - t(x,y)) \end{aligned} \tag{3}$$

There are three unknowns in this model, they are the propagation of an underwater medium $t(x,y)$, backscattered light B_∞, the radiant brightness of an object $J(x,y)$.

According to Eq. 3 we can get Eq. 4:

$$J(x,y) = \frac{I(x,y) - B_{\infty}}{t(x,y)} + B_{\infty} \tag{4}$$

The light field image is expressed as $LF(u,v,s,t)$. Based on this model, the formula of the restoration and enhancement of underwater light field image can be obtained.

Estimation of $t_{s,t}(u,v)$ for underwater light field images:

$$LF_{s,t}^{DC}(u,v) = \min_{pix \in \Omega(u,v)} \left(\min_{c \in \{r,g,b\}} \left(LF_{s,t}^{c}(u,v) \right) \right) \tag{5}$$

There is a relationship between $t_{s,t}(u,v)$ and the dark channel prior:

$$t_{s,t}(u,v) = 1 - LF_{s,t}^{DC}(u,v) = 1 - \min_{pix \in \Omega(u,v)} \left(\min_{c \in \{r,g,b\}} \left(LF_{s,t}^{c}(u,v) \right) \right) \tag{6}$$

In order to make $t_{s,t}(u,v) \in [0,\ 1]$,the input image $LF(u,v,s,t)$ needs to be transformed into a floating point image first and its range is $[0,\ 1]$.

Backscattering estimation in light field image is expressed as follow:

$$B_{\infty s,t}^{c}(u,v) = \max_{(u_*,v_*) \in M_{DC}^{\psi(u,v)}} \left(\min_{pix \in \Omega(u_*,v_*)} \left(LF_{s,t}^{c}(pix_u, pix_v) \right) \right) \tag{7}$$

The original image of the light field is expressed as $LF(u,v,s,t)$, and after restoration the image is expressed as $LF^{'}(u,v,s,t)$, through Eqs. 4–7 the underwater image haze removal formula can be expressed as follow:

$$LF^{'}(u,v,s,t) = \frac{LF(u,v,s,t) - B_{\infty s,t}^{c}(u,v)}{t_{s,t}(u,v)} + B_{\infty s,t}^{c}(u,v) \tag{8}$$

After establishing the underwater image restoration algorithm, the scale has an important influence on image enhancement. When the local area is large, the specular reflection and light effect can be effectively reduced, but the contrast effect is not ideal, there will still be atomized blur effects. On the contrary, the small area is very helpful to improve the contrast and haze removal. In this paper, we use the Pyramid image fusion method to fuse the light field image, the input images are

small-scale underwater restoration light field image, large-scale underwater restoration light field image and Laplace light field image. Each image corresponds to the corresponding weight. For images we use Laplace Pyramid and for weights we use the Pyramid of Gauss. Then the images and weights are fused to the corresponding layer image, the refused image is expressed as follow:

$$LF^{l}(u,v,s,t) = \sum_{k=1}^{K} \left\{ G^{l}(\bar{W}_k(u,v,s,t)) L^{l}(LF_k(u,v,s,t)) \right\} \tag{9}$$

2.2 Refocus and All-Focus

After the restoration and enhancement, there is still some local focus which means that some areas are fuzzy in the image. In order to get a clearer image, the fuzzy areas need to be refocused. The refocus formula is as follow:

$$\begin{bmatrix} u' \\ v' \\ s' \\ t' \end{bmatrix} = \begin{bmatrix} 1 & 0 & d & 0 \\ 0 & 1 & 0 & d \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ s \\ t \end{bmatrix}, \; T_c = \begin{bmatrix} 1 & 0 & d & 0 \\ 0 & 1 & 0 & d \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{10}$$

$$LF^{r}(u,v,s,t) = LF^{l}(u,v,s,t) T_c \tag{11}$$

where d is the distance between micro-lens array plane and Image sensor plane.

However, there are still some areas that are not clear even the image has been refocused. The reason is that when refocusing on other areas, the resolution of old ones is degenerated. In this situation, we can all-focus on the refocused images of different parameters to get the final clear image.

2.3 The Algorithm for Restoration and Enhancement

Based on the theory above, an algorithm is proposed for underwater light field image. In the restoration and enhancement, the image is converted into four-dimensional data first due to the properties of light field image. When getting the four-dimensional data, restore and enhance each angle image of the

four-dimensional data. The raw image is expressed as $LF(u, v, s, t)$, after the conversion, each angle image of four-dimensional data is expressed as $LF_{s,t}(u, v)$, restore and enhance $LF_{s,t}(u, v)$:

$$LF''_{s,t}(u, v) = \frac{LF_{s,t}(u, v) - B^{c}_{\infty s,t}(u, v)}{t_{s,t}(u, v)} + B^{c}_{\infty s,t}(u, v) \tag{12}$$

$$LF^{l'}_{s,t}(u, v) = \sum_{k=1}^{K} \left\{ G^{l}\left(\bar{W}^{k}_{s,t}(u, v)\right) L^{l}\left(LF^{k}_{s,t}(u, v)\right) \right\} \tag{13}$$

After the restoration and enhancement for each angle image of four-dimensional data, the images are converted into four-dimensional data $LF^{l'}(u, v, s, t)$ again. According to Eqs. 10 and 11, refocus on $LF^{l'}(u, v, s, t)$:

$$LF^{r'}(u, v, s, t) = LF^{l'}(u, v, s, t) T_c \tag{14}$$

When getting the refocused images of different parameters, the whole clear image can be obtained by all-focus.

3 Experiment

Using the Lytro light field camera as the experimental equipment in the course of the experiment, restore and enhance the images taken by this camera. In addition to the proposed algorithm, there is also another algorithm as a contrast. The principle of the algorithm is the same to the proposed one but the order of haze removal and the conversion to four-dimensional data is different. Here are the final results of these two algorithms.

3.1 Result of the Contrast Algorithm

For the contrast algorithm, the underwater light field image is restored and enhanced first. Then convert the image into four-dimensional data. When getting the data, refocus and all-focus on the four-dimensional data to get final clear image.

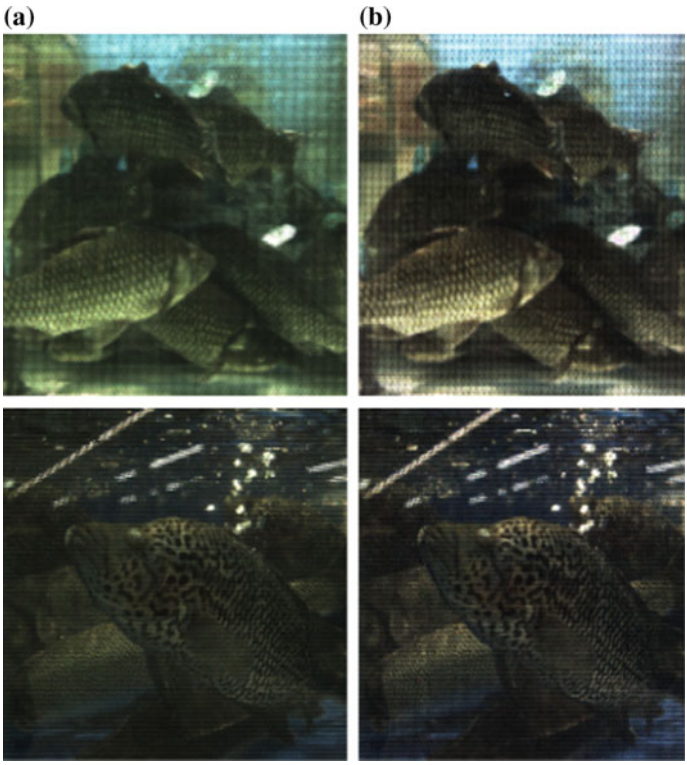


Fig. 1 Restored and enhanced results. **a** Original image. **b** Restored and enhanced image

We can see from Fig. 1 that the contrast algorithm has good effects on the restoration and enhancement. But some areas are still a little fuzzy that we can't get more details from the image. So the restored and enhanced image needs to be refocused to make these fuzzy areas clearer.

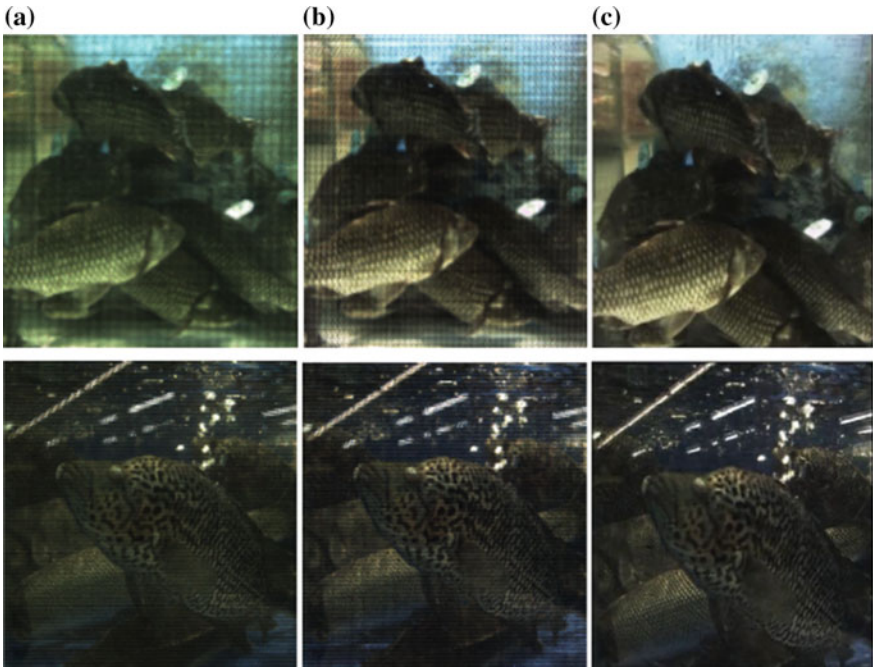


Fig. 2 Refocused results. **a** Original image. **b** Restored and enhanced image. **c** Refocused image

As Fig. 2 shows that the refocused image is clearer than the restored and enhanced image, some fuzzy areas in (b) become clear after being refocused. However, some old clear areas looks a little fuzzy in the image after being refocused, and this can be handled through all-focus. The experiment shows that the all-focused image is clearer in the whole image just like Fig. 3.

3.2 Result of the Proposed Algorithm

In the proposed algorithm, the light field image is converted into four-dimensional data first. When getting the four-dimensional data, restore and enhance for each angle image of the data.

From Fig. 4, we can see that the proposed algorithm has good effects on the restoration and enhancement. However, the same with the contrast algorithm, there are still some fuzzy areas. So the image needs to be refocused to get these areas clearer.

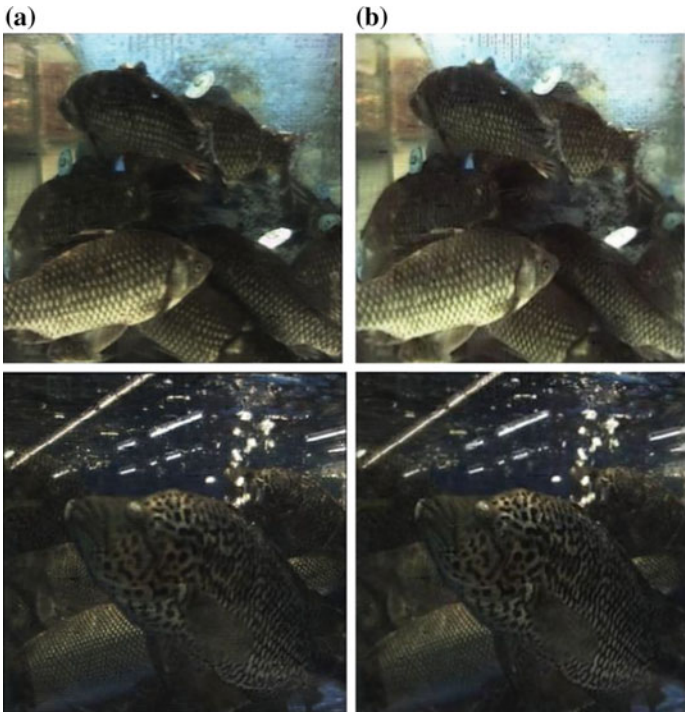


Fig. 3 **a** Refocused image. **b** All-focused image

After being refocused, we can see from Fig. 5 that the refocused image is clearer than the restored and enhanced image. Nevertheless, when refocusing on other areas, the old areas will become a little fuzzy. With this problem, it can be handled through all-focus. And the experiment shows that the all-focused image is clearer in the whole image just like Fig. 6.

3.3 Clarity Evaluation

These two algorithms both have good effects on restoration and enhancement of underwater light field image, but we do not know which one is better. So the clarity evaluation function is introduced to evaluate the final image. There are some

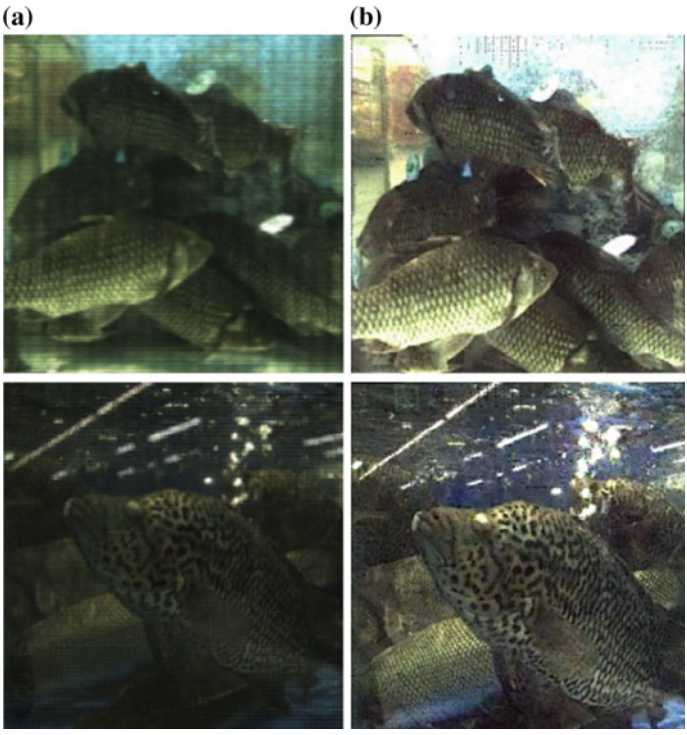


Fig. 4 **a** Original image. **b** Restored and enhanced image

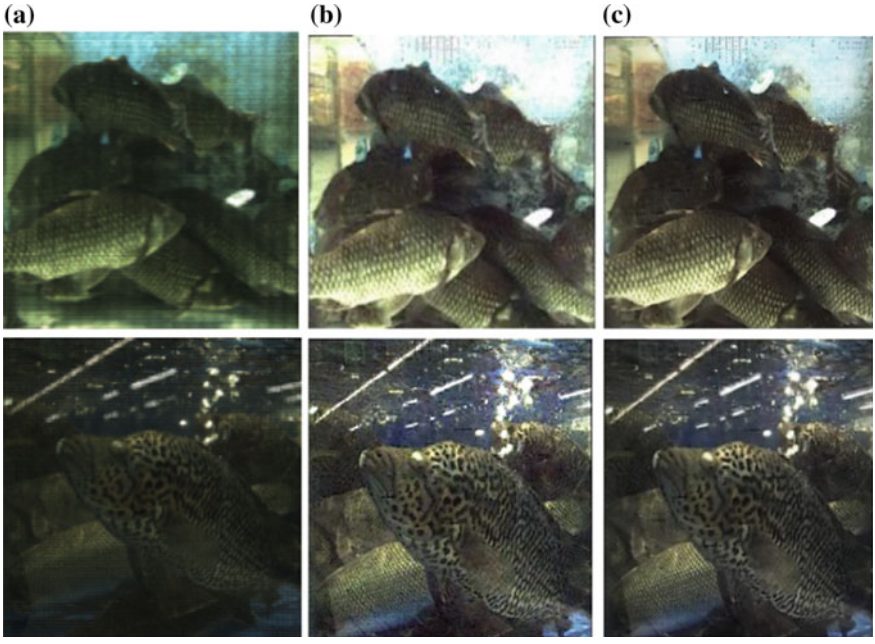


Fig. 5 **a** Original image. **b** Restored and enhanced image. **c** Refocused image

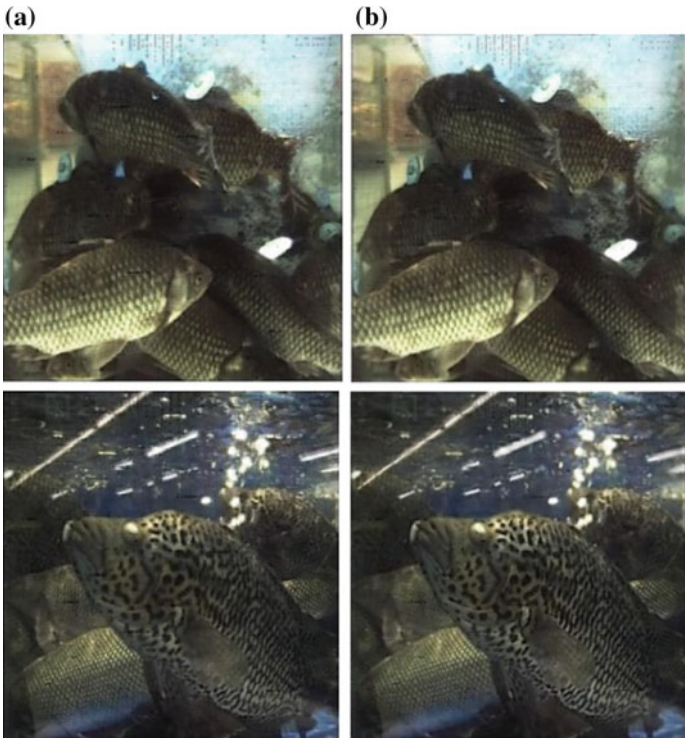


Fig. 6 **a** Refocused image. **b** All-focused image

common clarity evaluation functions such as Tenengrad gradient function, Laplacian gradient function, variance function, two step function and so on. In this paper we use variance function to evaluate the clarity of the restored and enhanced images obtained by different methods. Gray difference of image being focused clearly is larger than image being focused blurry, its variance is larger. So we can measure the clarity of the image through the variance of the image gray data, the greater the variance, the better the definition. In the evaluation, we choose the same parameter refocused image from two different algorithms, and then calculate the variance of the image gray data.

We can see from Fig. 7 that the refocused images of the proposed algorithm have a larger variance than images of the contrast algorithm, so the proposed algorithm is better in the restoration and enhancement of underwater light field image.

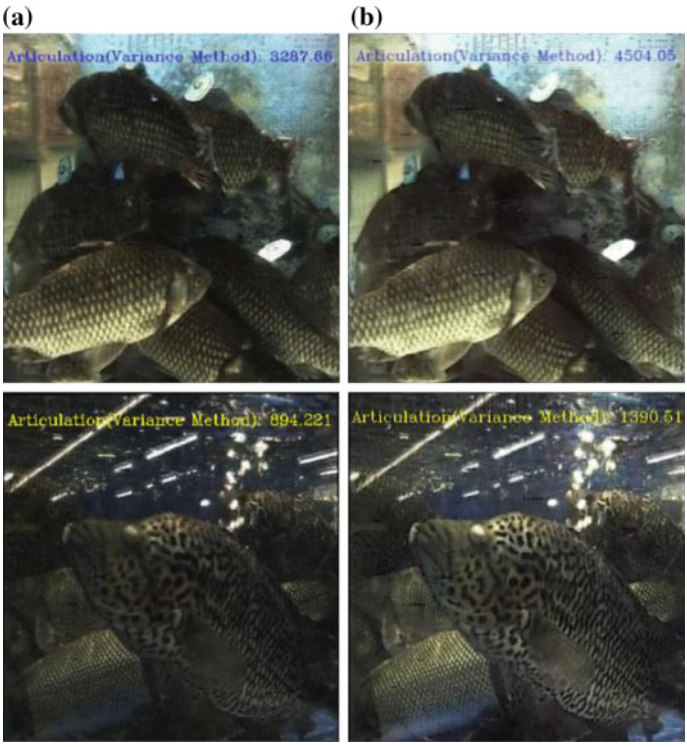


Fig. 7 **a** The variance of the image for the first algorithm. **b** The variance of the image for the second algorithm

4 Conclusion

From the experiment, the dark channel theory and Pyramid image fusion show good effects on the restoration and enhancement of underwater light field image. After haze removal and image fusion, however, there are still some fuzzy areas in the image. Under this circumstance, the image can be refocused and all-focused to become clearer. The light field image not only records the brightness of the spatial position, but also the direction of the light which means that it can be refocused on the image according to the information of the light field. Based on different distance parameters, images with diverse clear areas are obtained by being refocused. After the refocus, a whole clear image can be obtained by all-focus. The final result shows that images which are refocused and all-focused are clearer than images after haze removal and Pyramid image fusion.

Acknowledgements This research was partially supported by the National Nature Science Foundation of China (Grant no. 51575332 and no. 61673252) and the key research project of Ministry of science and technology (Grant no. 2016YFC0302401).

References

1. McGlamery BL (1980) A computer model for underwater camera systems. Ocean optics VI. Int Soc Optics Photonics, pp 221–231
2. Arnold-Bos A, Malkasset JP, Kervern G (2005) Towards a model-free denoising of underwater optical images. In: Arnold-Bos A (ed) Oceans 2005-Europe. Brest, France: IEEE Computer Society, pp 527–532
3. Arnold-Bos A, Malkasset JP, Kervern G (2005) A preprocessing framework for automatic underwater images denoising. In: Arnold-Bos A (ed) Proceedings of the european conference on propagation and systems 2005. Brest, France: European Conference on Propagation and Systems, pp 15–18
4. Chao L, Wang M (2010) Removal of water scattering. In: Chao L (ed) Proceedings of the 2010 international conference on computer engineering and technology. Chengdu, China: IEEE computer society 445 Hoes Lane-P. O. Box 1331 Piscataway NJ 08855–1331 United States, pp 235–239
5. Zhang X, Li C (2014) Calibration and imaging model of light field camera with microlens array. Acta Optica Sinica 34(12):1211005
6. Gortler SJ, Grzeszczuk R, Szeliski R et al (1996) The lumigraph. In: Proceedings of the 23rd annual conference on computer graphics and interactive techniques, pp 43–54
7. Mcmillan L, Bishop G (1996) Plenoptic modeling: an image-based rendering system. In: Proceedings of the 22nd annual conference on computer graphics and interactive techniques, pp 43–54
8. Nava FP, Luke JP (2009) Simultaneous estimation of super-resolved depth and all-in-focus images from a plenoptic camera. The true vision-capture, transmission and display of 3D video, pp 1–4
9. Ng R, Levoy M, Bredif M et al (2005) Light field photography with a hand—held plenoptic camera. Computer science technical report
10. He K, Sun J, Tang X (2009) Single image haze removal using dark channel prior. In: IEEE CVPR
11. Ancuti CO, Ancuti C, Bekaert P (2010) Effective single image dehazing by fusion. IEEE
12. Ancuti C, Ancuti CO (2016) Multi-scale underwater descattering. ICPR, pp 4202–4207
13. Levoy M, Hanrahan P (1996) Light field rendering. In: ACM trans proceedings of the 23rd annual conference on computer graphics and interactive techniques, pp 31–42
14. Zhou C, Nayar SK (2011) Computational cameras: convergence of optics and processing. IEEE Trans Image Process 20(12):3322–3340
15. Pertuz S, Puig D, Garcia MA, Fusiello A (2013) Generation of all-in-focus images by noise-robust selective fusion of limited depth-of-field images. IEEE Trans Image Process 22 (3):1242–1251
16. Tian J, Chen L (2012) Adaptive multi-focus image fusion using a wavelet-based statistical sharpness measure. Sig Process 92(9):2137–2146
17. Jhohura FT, Howlader T, Rahman SMM (2014) Bayesian fusion of ensemble of multifocused noisy images. Circuits Sys Signal Process 1–22
18. Pertuz S, Garcia MA, Puig D (2014) Efficient focus sampling through depth-of-field calibration. Int J Comput Vision 1–12
19. Teixeira R, Aizawa K (2014) Simultaneous acquisition of multiple images with higher dynamic range. In: 2014 IEEE International Conference on acoustics, speech and signal processing (ICASSP), pp 1355–1359. IEEE

Optimizing Leather Cutting Process in Make-to-Order Production to Increase Hide Utilization

Giuseppe Fragapane, Oladipupo Olaitan, Erlend Alfnes and Jan Ola Strandhagen

Abstract The growing trend of customization in the leather furniture industry has increased the need for many companies to switch from Make-To-Stock (MTS) to Make-To-Order (MTO) strategy. The shift to high-mix and low-volume presents new challenges for the production. This research presents concepts for the leather cutting process to increase the available calculation time for nesting and increase the selection for the nesting process, in order to maximize the hide utilization. The contribution to theory is a cutting process concept towards MTO in the leather industry.

Keywords Leather cutting process · Nesting · Make-to-order

1 Introduction

Mass customization is one of the leading strategies in satisfying customers' needs, and a competitive strategy in today's markets [1, 2]. The growing trend of mass customization can be seen in the leather industry [3] and it enables companies to switch from MTS to MTO strategy. The MTO strategy enables companies to produce high variety of products in small quantities and satisfies the unique needs of their customers [4]. Companies benefit from the MTO strategy in terms of cost and profit due to lower inventory levels, maximum sales, elimination of material waste, flexible production and customer satisfaction [5]. The shift from MTS to MTO strategy and moving the Costumer Order Decoupling Point (CODP) upstream presents new challenges for production in the leather industries.

For the leather cutting, the amount of different shapes to cut of the leather hide increase immensely while the batch size reduces significantly. The more different shapes there are available to nest, the more possible constellations of nesting them

G. Fragapane (✉) · O. Olaitan · E. Alfnes · J. O. Strandhagen
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, Norway
e-mail: giuseppe.fragapane@ntnu.no

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_10

on hides are possible. The number of possible constellation increases exponentially, and this calls for more process time to find the best fit of the parts on the hides. Generally, the more shapes there are to be nested, the better are the results for utilization and vice versa [6].

The efficient utilization of the leather hide has prime importance [7], which can be attributed to the fact that the cost of the leather hide often constitutes a significant proportion of the manufacturing costs, such that increasing its utilization provides significant cost savings [8, 9]. As a result, there has been a lot of attention in the leather industry to improve the nesting algorithms for determining how to fit variety of products onto hides [9–11]. Existing research has focused on developing algorithms that can determine the best possible ways to fit irregular product shapes and characteristics on leather hides, in order to achieve the best possible hide utilization within the shortest time possible [6, 9].

However, to most companies, the evaluation time and computational resources needed by these algorithms to deliver optimal results are not affordable [8]. This often leads to the acceptance of sub-optimal results for the hide utilization, and of results that do not suit pre-determined production plans.

This research investigates and presents possible methods to increase the time available for the nesting process in order to improve utilization. It analyses the interconnections between the manufacturing steps and proposes ways in which they can be re-organised. Its aim will be to address the question of how the nesting and cutting processes should be designed in order to maximize the hide utilization for high mix and low volume production.

The commonly used method will be compared with the suggested solution. The result of the investigation will stimulate further research to optimize leather cutting process in order to maximize the hide utilization.

2 Literature Review

2.1 Leather Nesting Problem

The roots of the Leather Nesting Problem derive from the cutting problem discussed in research since 1970s, in the metal, glass, wood, plastics, textiles and leather industries [10]. The cutting problem seeks to determine within given dimension the highest utilization of raw material through a heuristic approach [12]. The nesting problem is a two-dimensional cutting problem, which can occur in various production processes in leather industries [13]. Nesting is defined as "where more than one piece of irregular shape must be placed in a configuration with the other piece(s) in order to optimize an objective" [14]. The Leather Nesting Problem is one of the hardest two-dimensional cutting problems [9]. It allocates a set of irregular shapes within natural leather hides with highly irregular contours and heterogeneous interior due to holes and quality zones, in order to utilize most of the

leather hide [9, 15]. Traditionally the nesting process is performed manually and results in low nesting efficiency and utilization rate of the material [13]. Surveys showed that automated in comparison to manual nesting and cutting results in higher utilization and shorter process time [6].

Many researchers focus on optimizing algorithms for nesting to increase the hide utilization [8, 13, 16–18]. Because of the high cost of leather, small improvements in hide utilization have a great effect on profitability [7]. However, nesting cannot be seen as an isolated mathematical problem. Several processes are also affected by the nesting process. Nesting is connected closely to logistic planning and order management [19]. Factors like due dates, batch size and setup time are connected and have to be considered. Consequently, the nesting problem and the scheduling problem have to be addressed simultaneously. Surveys show that advanced scheduling techniques in the leather industry lead to significant improvements in utilization, idle time, make span and tardiness reduction [20]. Not taking the correlation between planning, scheduling and nesting into account will decrease the material utilization and overall production objectives and activities [21, 22].

2.2 *Leather Cutting Process*

Nesting is one of the essential steps in the leather cutting process, which consists of the steps of (1) Quality Control (2) Scanning (3) Nesting and (4) Cutting, as shown in Fig. 1.

Skilled workers have to set the hide on the machine and straighten the folds of the hide to a flat surface. In quality control, defects of the hides have to be identified and marked into zones. Because the range of defects is wide and various, automated defect detectors are still inefficient in identifying and evaluating the whole range of defects [23, 24]. Experienced workers can evaluate the defects best and assign them to five different quality categories of A, B, C, D and holes [9]. Scanning the hide reproduces an image with the contour of the hide. The image combined with the different quality zones is the basis for the nesting process. Shapes that will be nested on hides have different quality demands. For instance, the viewing and hardwearing areas of a leather furniture require the highest quality, whereas shapes with low visual areas can be allocated to lower quality areas. The nesting of the parts is a time intensive calculation operation due to the high number of possible

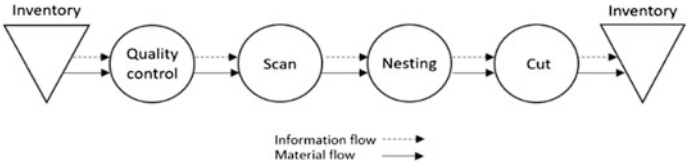


Fig. 1 Leather cutting process

constellations. The constellation with the highest hide utilization will be the template for cutting. The resulting template and, therefore, the contour of the parts are cut with a vibrating knife or a punching tool. In a last step, the cut parts are collected and forwarded to the next production process.

3 Process Reorganisation for Improved Leather Utilization

In this section, a concept is described for the reorganisation of the manufacturing steps, such that it enables to nest parts on several hides and decide with the support of an evaluation logic on the hide, which hide selection results in the highest utilization. The nesting process will be outside of the value adding time and thus more time can be dedicated to determining the best constellation.

As shown in Fig. 2, the new arrangement of the process is divided into three main processes of (1) Preparing and storing the data and material, (2) nesting, evaluating and deciding on a hide and (3) cutting. In the first step, hides are quality checked and scanned. These hides receive an identification number and are stored. The information of the marked quality zones and the scanned layout of each hide are stored in a logical database. The data of the hides enables in a second step to nest the parts from a pool of orders on several hides and evaluate the utilization. This evaluation can be supported by different decision logics. Different decision support logics can be designed and applied, depending on the company's preference and the production demands. This research will propose and investigate three different decision logics, as described in the following paragraphs, and illustrated in Fig. 3.

TO-BE 1: Select first hide, which reaches a minimum predefined percentage of hide utilization, for cutting. Parts are nested on different images of scanned hides. If a minimum predefined utilization, e.g. 70%, is passed, the template and hide are sent for cutting.

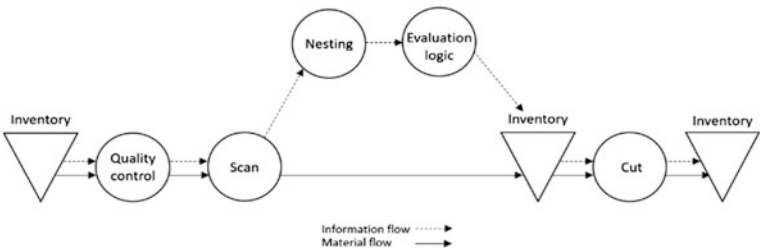


Fig. 2 Suggested leather cutting process reorganisation

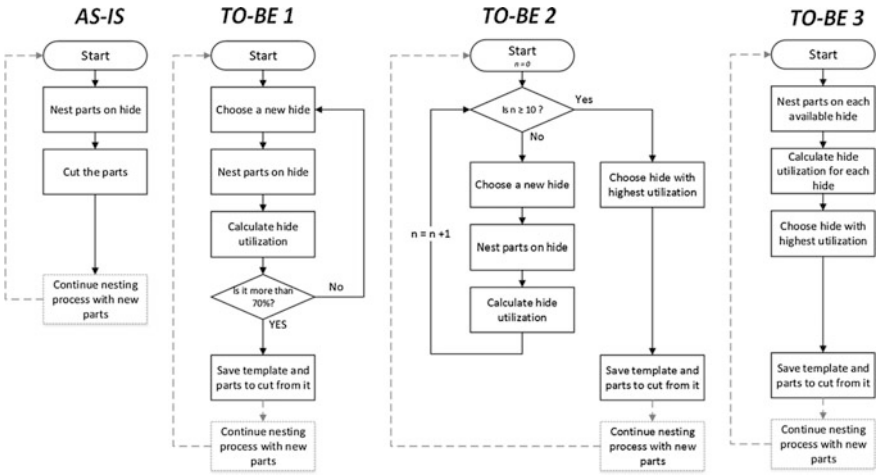


Fig. 3 Evaluation logic

TO-BE 2: Select hide, which reaches the highest hide utilization, within a defined range of hides for cutting. Parts are nested on a range of different scanned hide images, e.g. 10. The template and hide that achieves the highest hide utilization are sent for cutting.

TO-BE 3: Select and cut the hide that achieves the highest utilization among all available hides. Parts are nested on all images of scanned hides. The template and corresponding hide that achieved highest in hide utilization are sent for cutting.

In the last step, the hides are released from inventory and positioned on the cutting machines. The identification number of the hide is used to select the corresponding template which is then used to cut the parts out of the hide.

4 Discussion, Conclusion and Recommendation

The main strategy in the leather industry has traditionally been mass production and MTS strategy. Mass production typically has fewer product variants which again results in less variety of part shapes. This tradition has formed the current knowledge and understanding of the leather cutting processes. Figure 4 shows the variety and volume trade-off for the production strategies.

Low shape variety is unfavorable for hide utilization due to low possible nesting constellations. In MTS production, the batch size is kept large in order to use shapes several times in a nesting attempt. This way, the possible constellations and the leather hide utilization increase. Production planning can identify and define the batch sizes to meet short lead times and high hide utilization.

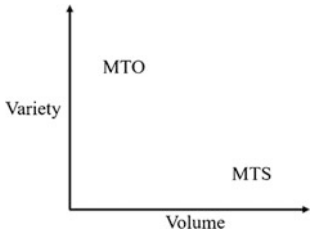


Fig. 4 Position of MTS and MTO in 'volume versus variety' based on [4]

The larger the batch size and the lower the batch variety are, the less important the variety of the hides is. Considering several hides within the nesting process and in a nesting attempt is of minor significance in mass production and MTS production.

The shift to mass customization and MTO strategy changes the value chain and makes it necessary to investigate and rearrange the processes. Since the leather parts are customized and the CODP moves upstream into leather cutting process and changes it from a forecast-driven to a customer-order driven process. The switch to customer-order driven process affects the batch size and parts variety significantly. The batch size decreases while the variety of the shapes to be cut increases. High variety of shapes results in more combinations and constellations in the nesting process and, consequently, more time is needed to calculate them. At the start of the nesting process, a wide choice-which is favorable-is available; however, towards the end of the batch and process, only less and complex parts are available. The utilization differ significantly between the beginning and end of a batch.

Combining similar orders to increase the batch size could lead to similar hide utilization results as in MTS production. Orders must be buffered and grouped. This will result in a high-mix production to an increase of throughput time and longer delivery dates. Long delivery dates reduce the competitiveness in today's markets. More flexibility is needed for a MTO production to hold a constant and high utilization.

This research presented a new arrangement of the cutting process and allocated the nesting process outside of the value added time to increase the available calculation time. This allows the nesting process to calculate more different constellations and combinations of the shapes onto several hides. Considering more hides in the process increases the probability of achieving higher hide utilization. Complex shapes and hides with poor quality can be used in several nesting attempts and so increase the chances of achieving higher utilization. This arrangement can take in consideration small batches and high variety.

The evaluation logic can be adjusted to the production demands. Each evaluation logic influences the hide utilization, but also the inventories and process time. The amount of hides taken in account in the evaluation logic will define the size of the inventory and the time needed for nesting. The more strings are included in the evaluation logic, the more resources, for instance inventory space, IT support or servers, will be needed. The ranking of different evaluation logics can be seen in Table 1.

Table 1 Ranking of logics

	Hide utilization	Process time	Inventories
TO-BE 1	3	2	2
TO-BE 2	2	1	1
TO-BE 3	1	3	3

Comparison of logic with each other (1 = best, 3 = worst)

TO-BE 1 is ideal, if the hide utilization has to achieve a minimum and be constant. The challenge will be to determine the minimum hide utilization. Setting the limit low, satisfactory results can be expected in a short calculation time. Setting the limit high can result in an endless process.

TO-BE 2 finds the best hide utilization within a range of hides. Determining the range of hide for evaluation will define process time and needed inventory. Both will be constant which is favorable for planning and scheduling.

Considering all available hides in the nesting process is challenging due to the nesting process time. TO-BE 3 can be applied, if the benefit of the hide utilization justifies the long process time. This can be the case for low volume and exceptionally high raw material costs.

The new arrangement of the process and the proposed evaluation logic will significantly improve hide utilization. However, these changes imply more and new process steps and more handling of the hide. These changes can be justified, if the higher utilization of the hide and resulting savings would cover the additional costs.

Future research will be aimed at further developing the evaluation logics, in order to keep the process time and inventories low.

Acknowledgements This research has been carried out as part of SØM4.0 project, as funded by the Norwegian Research Council's BIA programme. The authors wish to thank the project partners for facilitating this work.

References

1. Daaboul J et al (2011) Design for mass customization: Product variety vs. process variety. CIRP Ann Manuf Technol 60(1):169–174
2. Da Silveira G, Borenstein D, Fogliatto FS (2001) Mass customization: literature review and research directions. Int J Prod Econ 72(1):1–13
3. Barnett L, Rahimifard S, Newman S (2004) Distributed scheduling to support mass customization in the shoe industry. Int J Comput Integr Manuf 17(7):623–632
4. Stevenson M, Hendry LC, Kingsman BG (2005) A review of production planning and control: the applicability of key concepts to the make-to-order industry. Int J Prod Res 43 (5):869–898
5. Pollard D, Chuo S, Lee B (2016) Strategies for mass customization. J Bus Econ Res (Online) 14(3):101
6. Elamvazuthi I, Kamaruddin S, Azmi MS (2008) Automation of nesting and cutting processes of leather furniture production: a case study. Int J Mech Mech Eng 9(10):25–29

7. Crispin A et al (2005) Genetic algorithm coding methods for leather nesting. Appl Intell 23 (1):9–20
8. Baldacci R et al (2014) Algorithms for nesting with defects. Discrete Appl Math 163:17–33
9. Alves C et al (2012) New constructive algorithms for leather nesting in the automotive industry. Comput Oper Res 39(7):1487–1505
10. Dyckhoff H (1990) A typology of cutting and packing problems. Eur J Oper Res 44(2):145–159
11. Elkeran A (2013) A new approach for sheet nesting problem using guided cuckoo search and pairwise clustering. Eur J Oper Res 231(3):757–769
12. Dowsland KA, Dowsland WB (1992) Packing problems. Eur J Oper Res 56(1):2–14
13. Huang Y et al (2005) Research on an intelligent leather nesting system. In: ICMIT 2005: merchatronics, MEMS, and smart materials. International Society for Optics and Photonics
14. Bennell JA, Oliveira JF (2008) The geometry of nesting problems: a tutorial. Eur J Oper Res 184(2):397–415
15. Crispin AJ, Cheng K (2008) Backtracking greedy algorithm for cutting stock problems. In: Applied mechanics and materials. Trans Tech Publications
16. Yuping Z, Shouwei J, Chunli Z (2005) A very fast simulated re-annealing algorithm for the leather nesting problem. Int J Adv Manuf Technol 25(11–12):1113–1118
17. Uday A, Goodman ED, Debnath AA (2001) Nesting of irregular shapes using feature matching and parallel genetic algorithms. In: Citeseer
18. Heistermann J, Lengauer T (1995) The nesting problem in the leather manufacturing industry. Ann Oper Res 57(1):147–173
19. De Vin L, De Vries J, Streppel T (2000) Process planning for small batch manufacturing of sheet metal parts. Int J Prod Res 38(17):4273–4283
20. Habib A, Jilcha K, Berhan E(2015) Performance improvement by scheduling techniques: a case of leather industry development institute. In: Afro-European conference for industrial advancement. Springer
21. Chryssolouris G, Papakostas N, Mourtzis D (2000) A decision-making approach for nesting scheduling: a textile case. Int J Prod Res 38(17):4555–4564
22. Sakaguchi T et al (2012) A scheduling method with considering nesting for sheet metal processing. In: ASME/ISCIE 2012 international symposium on flexible automation. American Society of Mechanical Engineers
23. Chiou Y-C (2010) Intelligent segmentation method for real-time defect inspection system. Comput Ind 61(7):646–658
24. Yeh C, Perng D-B (2005) A reference standard of defect compensation for leather transactions. Int J Adv Manuf Technol 25(11):1197–1204

A Study on Material Properties of Sealed Rubber Cylinder for Compressible Packer

Lixin Lu, Yue Qian, Guiqin Li and Peter Mitrouchev

Abstract The stress–strain curve of the packer seal rubber cylinder material is obtained by uniaxial tensile test in this paper, and the material constants of the Yeoh model is fitted in Origin. During the test, two kinds of samples are tested, the side and intermediate cylinder, and two sets of figures are measured for each sample, totally four sets of data are obtained. By the fitting results of the side and middle cylinder are achieved by comparing the material constants for each sample respectively.

Keywords Packer · Sealed cylinder · Sealing performance · Strain energy function · Material constants

1 Introduction

The packer rubber cylinder is characterized by the hyperelasticity, when the force is very small, it can produce a great deformation. In the case of large strain, the hardening or the softening of rubber cylinder is accelerated, it has geometrical and material nonlinearity [1]. In the finite element analysis, the constitutive model of the rubber material is selected reasonably to obtain material constants.

In 2010, Chen Jian analyzes the impact of the tube material on the sealing performance by simulation and real test. The results show that the Oxidation phenomenon can reduce the comprehensive performance of the plastic cylinder, and eventually lead to the sealing failure [2]; In 2012, Yu Guijie uses the uniaxial tensile test of the rubber with the filler added, and the material constants of the

L. Lu · Y. Qian · G. Li (✉)
Shanghai Key Laboratory of Intelligent Manufacturing and Robotics, Shanghai University, Shanghai, China
e-mail: leeching@shu.edu.cn

P. Mitrouchev (✉)
University Grenoble Alpes, G-SCOP, 38031 Grenoble, France
e-mail: peter.mitrouchev@g-scop.inpg.fr

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_11

Mooney–Rivlin model and the Yeoh model are obtained. Finally, the sealing effect of rubber box with different elastic modulus is obtained by ANSYS [3]. Li Pengfei experiments on the three-rubber-tube packer, the stress–strain relationship of rubber material is obtained [4].

The Yeoh model is used to describe the mechanical properties. The mechanical experiments of the sealed cylinder are carried out, and then the material constants of the strain energy function are fitted.

2 Mathematical Model for Rubber Material

Yeoh strain energy function can be used to express the mechanical properties of the rubber material in large deformation, and is suitable for the pressure strain calculation model of a packer seal cylinder [5]. The material constants of Yeoh model can be determined by uniaxial tensile experiments [6].

The strain energy function can be considered as:

$$W = \sum_{ijk=0}^{n} C_{ijk}(I_1 - 3)^i (I_2 - 3)^j (I_3 - 3)^k \tag{1}$$

where C represents the material constant, invariants I_1, I_2, I_3 are defined as:

$$\begin{cases} I_1 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2 \\ I_2 = (\lambda_1\lambda_2)^2 + (\lambda_2\lambda_3)^2 + (\lambda_3\lambda_1)^2 \\ I_3 = (\lambda_1\lambda_2\lambda_3)^2 \end{cases} \tag{2}$$

Where λ_i is the main elongation ratio. Since the rubber cylinder is approximately incompressible, we consider $I_3 = 1$, (1) can be simplified as:

$$W = \sum_{ij=0}^{n} C_{ij}(I_1 - 3)^i (I_2 - 3)^j \tag{3}$$

The strain energy function in (3) can be written as:

$$W = C_{10}(I_1 - 3) + C_{20}(I_1 - 3)^2 + C_{30}(I_1 - 3)^3 \tag{4}$$

The relationship between the Kirchhoff stress tensor and the Green strain tensor is

$$t^{ij} = \frac{\partial W}{\partial I_1}\frac{\partial I_1}{\partial \gamma_{ij}} + \frac{\partial W}{\partial I_2}\frac{\partial I_2}{\partial \gamma_{ij}} + \frac{\partial W}{\partial I_3}\frac{\partial I_3}{\partial \gamma_{ij}} \tag{5}$$

Where t^{ij}, γ_{ij} respectively represents the Kirchhoff stress tensor and the Green strain tensor.According to the relationship between the stress tensor and the strain tensor,

$$\begin{cases} t_1 = 2\lambda_1\left[\frac{\partial W}{\partial I_1} + (\lambda_2^2 + \lambda_3^2)\frac{\partial W}{\partial I_2} + \lambda_2^2\lambda_3^2\frac{\partial W}{\partial I_3}\right] \\ t_2 = 2\lambda_2\left[\frac{\partial W}{\partial I_1} + (\lambda_1^2 + \lambda_3^2)\frac{\partial W}{\partial I_2} + \lambda_1^2\lambda_3^2\frac{\partial W}{\partial I_3}\right] \\ t_3 = 2\lambda_3\left[\frac{\partial W}{\partial I_1} + (\lambda_1^2 + \lambda_2^2)\frac{\partial W}{\partial I_2} + \lambda_1^2\lambda_2^2\frac{\partial W}{\partial I_3}\right] \end{cases} \tag{6}$$

Using uniaxial tensile tests,

$$\begin{cases} \sigma_2 = \sigma_3 = 0 \\ \lambda_2^2 = \lambda_3^2 = \dfrac{1}{\lambda_1} \end{cases} \tag{7}$$

Where t_1 is stress value, λ_1 is strain value, thus

$$Y = 2C_{10} + 4C_{20}X + 6C_{30}X^2 \tag{8}$$

In order to fit mechanical constants C_{10}, C_{20}, C_{30}, uniaxial tensile tests are carried on.

3 Experiment and Result

The uniaxial tensile test was carried out to obtain the stress and strain data of the packer seal cylinder. The size, trimming and measuring tools of the sample meet the requirements of the national standard GB /T 528-2009.

A cutter, complying with the requirements of GB /T2941, is used to cut the packer sealing cylinder, and the sample is measured by the thickness gauge. The cut sample is shown in Fig. 1.

SCT-57 (232) microcomputer control electronic testing machine is used in this test, as shown in Figs. 2 and 3.

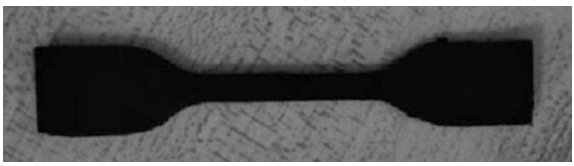


Fig. 1 The cut sample

Fig. 2 The testing machine

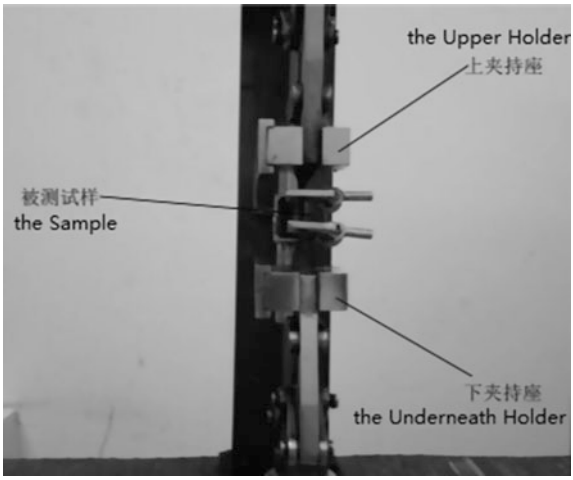


Fig. 3 The ongoing experiment

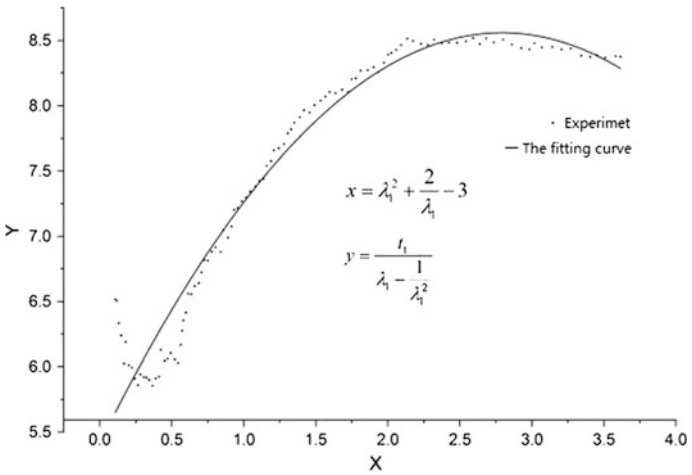


Fig. 4 The 1st curve of the intermediate cylinder

Two kinds of samples are tested, the side and intermediate cylinder, and two sets of figures are measured for each sample, totally 4 sets of data are obtained. The material constants of the Yeoh are fitted in Origin.

The two sets of stress and strain data are shown in Table 1 and Table 2.

According to the two sets of stress and strain data, the fitted curves of the intermediate cylinder are shown in Figs. 4 and 5.

From the data tables and fitting curves above, it can be seen that the data of the side rubber tube in two tables are basically the same, and the fitting results of the intermediate cylinder are also the same.

The error of the first group of data is less than the second group of data, so the fitting result of the first group of data is taken as the material constant of the side cylinder. The same to the intermediate cylinder, by comparing with two sets of figures, the fitting result of the first group of data is selected as the material constant. The final fitting results of the material constants of the side and the intermediate cylinder are shown in Table 3.

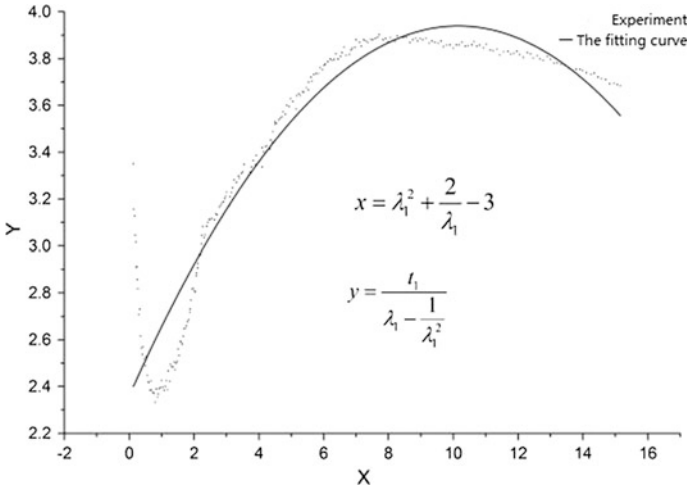


Fig. 5 The 2nd curve of the intermediate cylinder

Table 1 The first set of stress and strain data of the side cylinder

Number	Load/kN	Deformation/mm	Stress/MPa	Strain
1	0.00437	0.16936	0.54625	0.006774
2	0.00700	0.19053	0.87500	0.007621
3	0.00956	0.31755	1.19500	0.012702
4	0.01094	0.57159	1.36750	0.022864
5	0.01206	0.91031	1.50750	0.036412
•	•	•	•	•
•	•	•	•	•
•	•	•	•	•
24	0.03506	7.45184	4.38250	0.298074
25	0.03593	7.83290	4.49125	0.313316
26	0.03780	8.29864	4.72500	0.331946
27	0.03880	8.53151	4.85000	0.341260
28	0.03999	8.93374	4.99875	0.357350
•	•	•	•	•
•	•	•	•	•
•	•	•	•	•
119	0.16278	36.56059	20.34750	1.462424
120	0.16334	37.02633	20.41750	1.481053

Table 2 The second set of stress and strain data of the side cylinder

Number	Load/kN	Deformation/mm	Stress/MPa	Strain
1	0	0	0	0
2	0.00825	0.02117	1.03125	0.000847
3	0.01025	0.04234	1.28125	0.001694
4	0.01025	0.10585	1.56250	0.004234
5	0.01362	0.42340	1.70250	0.016936
•	•	•	•	•
•	•	•	•	•
•	•	•	•	•
24	0.03362	7.40950	4.20250	0.296380
25	0.03437	7.74822	4.29625	0.309929
26	0.03549	8.15045	4.43625	0.326018
27	0.03681	8.38332	4.60125	0.335333
28	0.03793	8.74321	4.74125	0.349728
•	•	•	•	•
•	•	•	•	•
•	•	•	•	•
119	0.14760	34.52827	18.45000	1.381130
120	0.14903	34.93050	18.62875	1.397220

Table 3 The fitting results of the side and intermediate cylinder

	Fitting coefficient	Value	Error		Fitting coefficient	Value	Error
The Side cylinder	C_{10}	5.562	0.045	The Intermediate cylinder	C_{10}	2.361	0.025
	C_{20}	2.097	0.053		C_{20}	0.311	0.009
	C_{30}	−0.319	0.013		C_{30}	−0.015	0.001

4 Conclusion

The stress and strain data of the sealed cylinder were obtained by uniaxial tensile test of the packer seal cylinder according to GB / T 528-2009. The material constants of the side rubber tube and the middle cylinder in Yeoh function are obtained by fitting the 4 sets of stress and strain data to facilitate the calculation of the sealing performance of the packer.

Acknowledgements This research is partly supported by the research program of Shanghai Science and technology committee (NO. 14DZ1204203).

References

1. Khajehsaeid H, Arghavani J, Naghdabadi R (2013) A hyperelastic constitutive model for rubber-like materials. Eur J Mechanics-A/Solids 38:144–151
2. Jian Chen, Boyuan Tian et al (2010) Failure Analysis of Compressible Packer rubber cylinder. Sci Technol Inf 31(11):80–81
3. Guijie Yu, Changjiang Li et al (2015) Sealing performance analysis on shaped rubber of compressed packers. Lubr Eng 40(1):45–46
4. Pengfei Li, Liu Xu et al (2016) Analysis of sealing performance of three-rubber-cylinder packer. Mech Manuf 622(54):51–52
5. Karimi A, Navidbakhsh M, Beigzadeh B et al (2013) Hyperelastic mechanical behavior of rat brain infected by plasmodium berghei experimental testing and constitutive modeling. Int J Damage Mech 2:1056
6. Faturechi R, Karimi A, Hashemi SA et al (2014) Mechanical characterization of peritoneum/fascia under uniaxial loading. J Biomater Tissue Eng 4(3):189–193

Finite Element Analysis of Sealing Performance for Multi Cylinder Packer

Li-xin Lu, Tao Deng, Gui-qin Li and Peter Mitrouchev

Abstract Sealing performance of multi cylinder packer is analyzed by using finite element analysis method in this paper. The setting process is analyzed by apply different sealing load to the multi cylinder packer, and it is found that the different sealing load will cause the different gap between the rubber tube and the central tube in the packer. Thus, the position of the main seal in the multi cylinder packer is thus obtained.

Keywords The finite element analysis · Multi cylinder packer
Seal · Sealing load

1 Introduction

Most of the deep and ultra deep wells are facing more severe geological conditions [1, 2]. Irregular topography, high temperature and high pressure downhole environment make the packer seal failure accident frequent [3, 4].

In the main components of the packer, compressive packer rubber is very important, which directly affects the function of the packer. Once the glue cylinder sealing effect is not very good, it will make it appear certain security risks, and even cause some economic losses. The statistical data of 2015 in Shengli Oilfield of

L. Lu · T. Deng · G. Li (✉)
Shanghai Key Laboratory of Intelligent Manufacturing and Robotics,
Shanghai University, Shanghai 200072, China
e-mail: leeching@shu.edu.cn

P. Mitrouchev (✉)
University Grenoble Alpes, G-SCOP, 38031 Grenoble, France
e-mail: peter.mitrouchev@g-scop.inpg.fr

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_12

Zhuangxi Oil production plant shows that the packer sealing failure cases each year up to more than ten, the problem lies in the contact pressure is too much concentrated in a single packer rubber [5]. In view of the problem of the sealing performance of the multi cylinder packer, Zhang Baoling puts forward the improvement method of the packer under high pressure [6], Zhang Debiao points out that the device can not only improve the sealing of the packer, but also increase the service life of the rubber cylinder, Yang Zhipeng, et al. pointed out that the umbrella type anti protrusion device can slow down the peristaltic relaxation of the sealant cylinder by the indoor test of packer rubber tube [7]. The method that analyze the sealing performance of the packer by experimental is not only expensive and time-consuming, but also some external factors can affect the experimental results. The finite element method is used to analyze the sealing of it, which can shorten the period of design and analysis, reduce the test time and cost, and improve the reliability of products and projects. Therefore, this paper try to through the finite element numerical simulation analyze the sealing performance of multi cylinder packer.

2 Establishment of Finite Element Model

In the analysis of the model by ANASYS, in order to save computational resources, we need to simplify the complex computing model. However, the results will inevitably affect the calculation accuracy. Considered that ANASY can be seamlessly connected with 3D design and improve the precision of calculation. The whole model is introduced into the finite element analysis software for analysis and calculation. In the SolidWorks, model will be saved as an IGES file, and import the ANASY software, then set the modal material parameters. Material is structural steel and the material constant of sealing cylinder is obtained by fitting the Yeoh model. The mechanical parameters of the material can be obtained by referring to the mechanical design manual [8] and the elastic modulus is 20,600 MPa and the Poisson's ratio is 0.25.

The mesh of the sealing cylinder is divided by polygon mapping mesh, and other non key parts such as the central oil pipe, casing and base are divided by free mesh. In the control of grid size, when the element size is less than 1 mm, the computation time is greatly increased, and the calculation results are basically unchanged, therefore, the element of size is 1 mm, the number of nodes is 1766, the number of cells is about 511, and the average value of unit mass is about 0.9424. Because the casing is provided with cement cementation, the casing is set to be fixed, and the central oil pipe, the pressing block and the base are provided with a radial displacement constraint, base and centre tubing binding.

3 Analytical Methods and Procedures

The 3 rubber packer is more and more widely applied in the engineering, its structure is shown in Fig. 1. When analyse and calculate the 3 rubber packer in ANASY, firstly, the load of 10–25 MPa was applied to the 3 rubber packer, and the sealing process was analyzed.

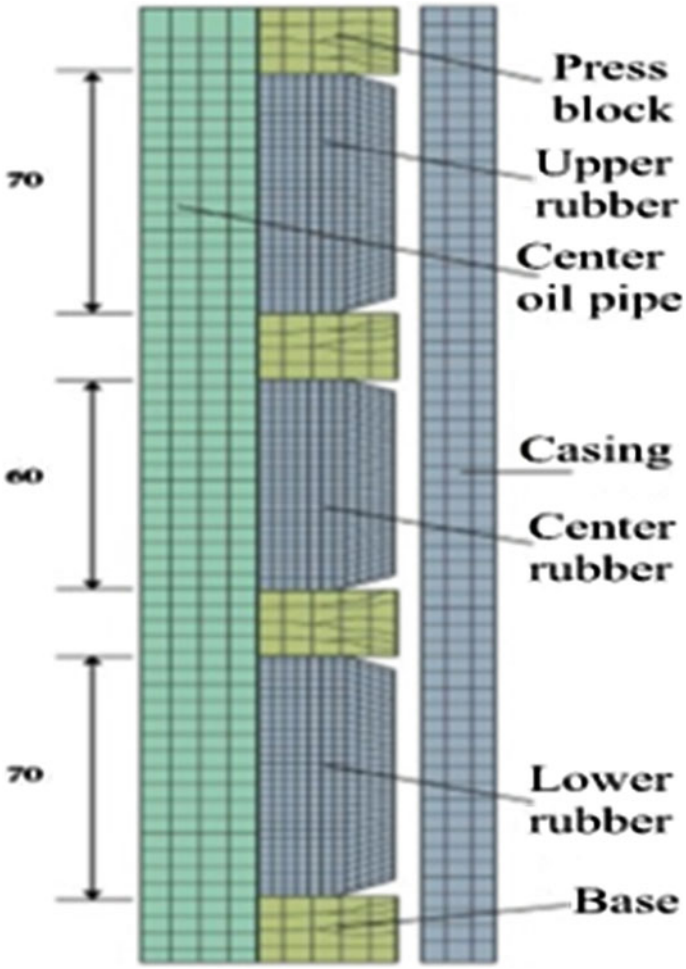


Fig. 1 Mesh of sealing structure

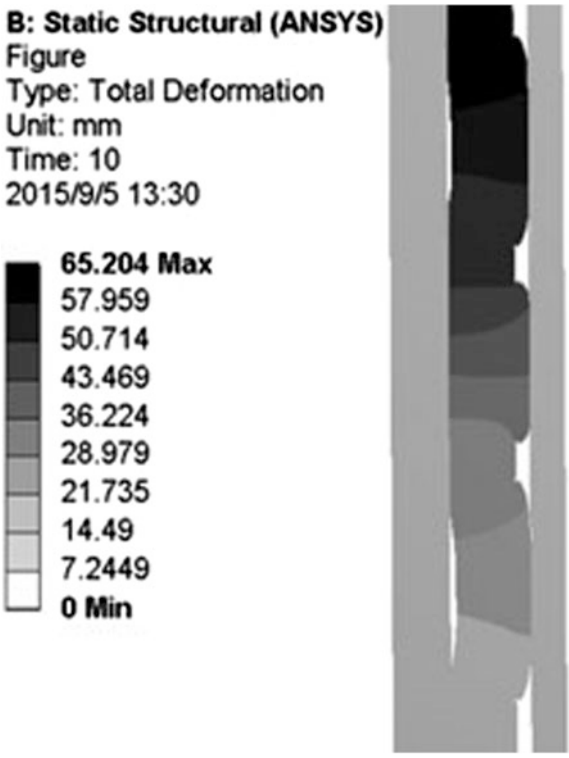


Fig. 2 Load 10 Mpa

The deformation of the sealing cylinder is shown in Fig. 2 under 10 MPa and Fig. 3 under 14 MPa load. As can be seen from Fig. 2, under the load is 10 MPa, the total deformation amount of the rubber cylinder is 65.204 mm, and the 3 rubber cylinders are initial contact with the inner wall of the casing, but there is a big gap between the upper and the lower cylinder. When the sealing load is increased to 14 MPa, as can be seen from Fig. 3, the total compression amount of the rubber cylinder is increased to 75.003 mm, and the clearance between the upper rubber tube and the central oil pipe is basically eliminated, but there is still a gap between the lower rubber tube and the central tubing, so it is necessary to increase the sealing load so as to increase the compression capacity of the lower tube.

Figures 4 and 5 are the amount of deformation of the 3 sealing cylinders under the sealing load is increased to 18 and 25 MPa. As can be seen from Fig. 4, when the sealing load is increased to 18 MPa, the total compression distance is 81.433 mm, but there is still a certain amount of clearance between the lower rubber tube and the central tubing, and the sealing load still needs to be improved.

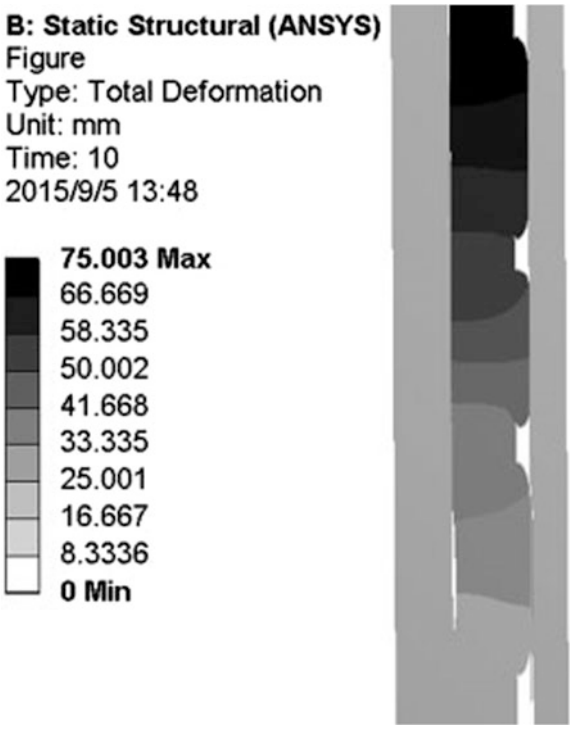


Fig. 3 Load 14 Mpa

As can be seen from Fig. 5, under the load is 25 MPa, the compression distance is increased to 90.173 mm, and the clearance between the upper rubber cylinder, the middle rubber cylinder, the lower rubber cylinder and the central oil pipe is eliminated, and the 3 rubber cylinders are fully compressed. Compared with the single plastic tube structure, the sealing load required for the full closure of the 3 plastic tubes increased from 15 to about 25 MPa. Therefore, in the design of the 3 packer, the sealing load should be appropriately increased to ensure that the 3 rubber cylinders can be fully compressed, so that the packer can be fully seated.

As can be seen from Fig. 6, the contact stress between the intermediate rubber cylinder and the inner wall of the casing is the highest and the intermediate rubber cylinder acts as the main sealing function. When the load is 10 and 14 MPa, the contact stress of the 3 rubber cylinders is approximately the same. This is because the 3 rubber cylinders are not fully compressed. With the increase of sealing load, when sealing load increased to 25 MPa, the maximum contact stress between the intermediate rubber cylinder and casing is about 7 MPa, while the maximum contact stress of the rubber cylinder is only about 3 MPa, so the intermediate rubber cylinder is fully compressed and has the highest contact stress.

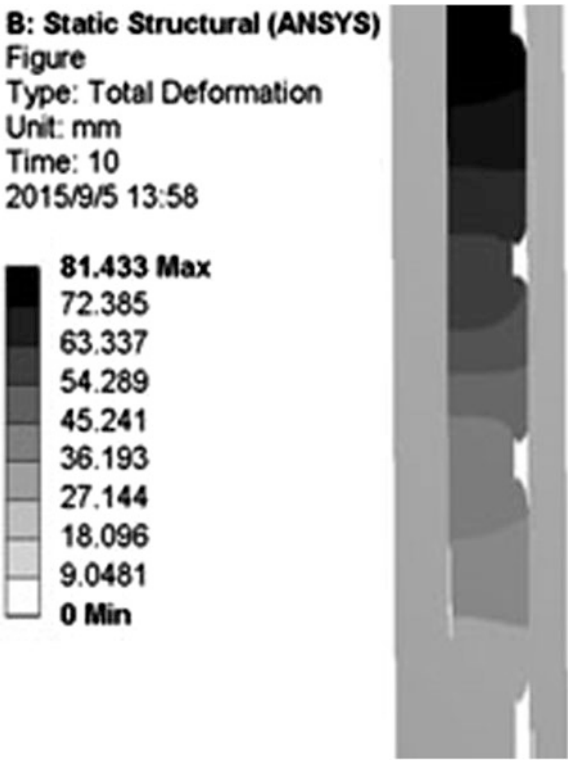


Fig. 4 Load 18 MPa

4 Conclusion

In this paper, the sealing structure model of the 3 rubber cylinder is established according to the actual size and saved as IGES file, and imported into ANASY software for material settings, then the mesh of the sealing cylinder is divided by polygon mapping mesh, and other non key parts such as the central oil pipe, casing and base are divided by free mesh. The bushing is fixed, the centre pipe, the pressing block and the base are provided with radial displacement restraint, and the base is bound with the central oil pipe. Finally, in the ANSYS, the sealing load of 10–25 MPa is applied to the 3 rubber packer. The results of simulation show that: in the sealing structure of the 3 rubber cylinder, the middle rubber cylinder plays a main role in sealing.

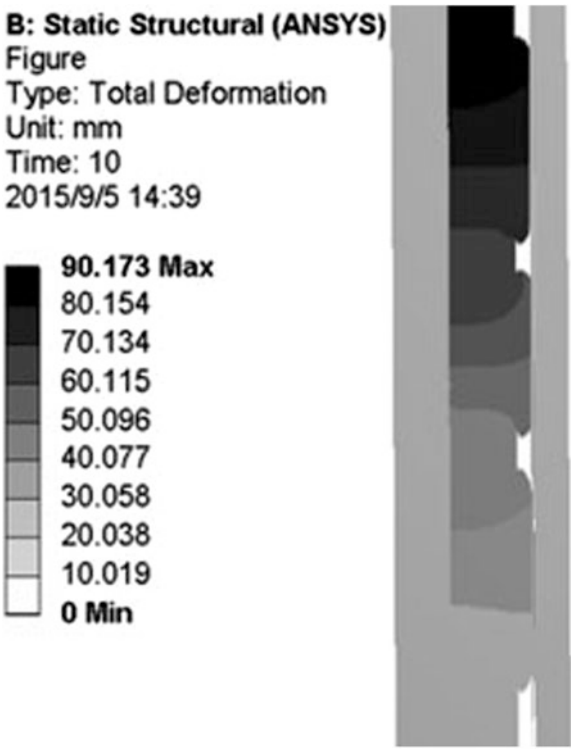


Fig. 5 Load 25 MPa

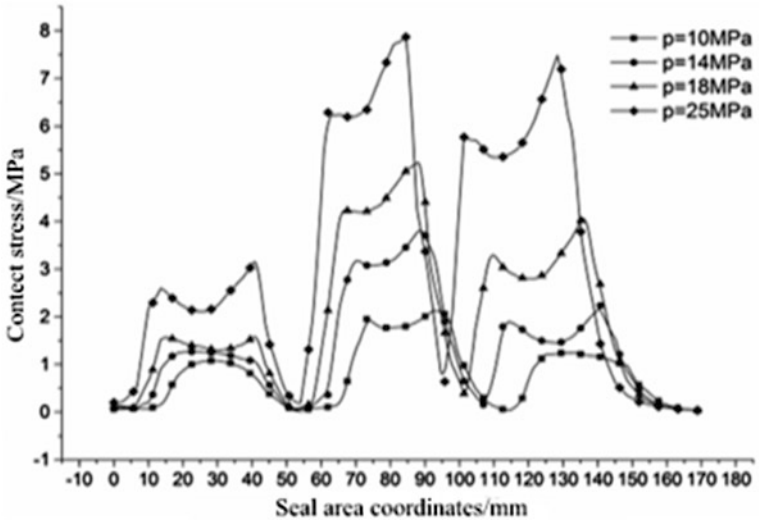


Fig. 6 Contact stress distribution

Acknowledgements This research is partly supported by the research programmer of Shanghai Science and technology committee (NO. 14DZ1204203).

References

1. Xinke Yang, Shaer Hai (2012) Discussion on design idea of rubber cylinder for compression packer. J Xinjiang Oil Gas 01:115–118
2. Dou Y, Ma Z, Chen J et al (2015) Finite element analysis of mechanical behavior of THT packer rubber cylinder. J Oil Gas Well Test 03:4-7-75
3. Zhang L, Shen G, Li Y et al (2007) Development and application of packer technology in China. J Petrol Mach 08:58-60-74-80
4. Zhang Yuhang Xu, Xiaobing Yang Ya (2015) Analysis of the development of packers at home and abroad. J Mech Eng 01:37–39
5. Qu Baolong Xu, Huajing Ma Wei (2015) Optimum design of packer rubber cylinder based on friction coefficient. J Petrol Mach 02:83–87
6. Zhang B, Wang X, Xu X (2009) Improvement of sealant cylinder for high pressure packer. J Petrol Field Mach 01:85–87
7. Yang Z, Liu D, Xu Y et al (2012) Experimental study on deformation of compression rubber cylinder shoulder. J Lubr Sealing 02:103–105
8. Datong Qin, Liyang Xie (2011) Manual of modern mechanical design, M. Chem Ind Press, Beijing

Experimental Study on the Effect of Surface Texture on the Dynamic Performance of Journal Bearing

Jin Zhang, Guoping Li, Xiaojing Wang, Xiong Xin and Zikai Hua

Abstract Growing attention has been paid to surface texture owing to its significance on bearing lubrication, load capacity and wear resistance. In this paper, in order to learn more about the dynamic performance of journal bearings and the vibration of bearing housing, a comparative experimental study between textured surface and non-textured surface journal bearing configurations is presented. The research was mainly divided into two parts. Firstly, the feature information of displacement under a range of loads from 0 N to 2500 N was obtained. Experimental measurements data of stiffness coefficients and damping coefficients under different static loads were calculated and analyzed respectively between the surface-textured and non-textured journal bearings. It can be learned that the values of fluid film stiffness coefficients and damping coefficients are increased via surface texture. Secondly, the vibration amplitude of bearing housing was measured. The experimental results show that appropriate texture distribution can certainly have better vibration damping effects according to acceleration amplitude. There is a good correlation under the relevant operating conditions compared with a theoretical model. The experimental results show that the appropriate texture distributed journal bearing provides a great improvement on stability than the non-textured journal bearing.

Keywords Surface texture · Dynamic performance · Journal bearing Vibration

J. Zhang · X. Wang (✉) · X. Xin · Z. Hua
Department of Mechatronic Engineering and Automation,
Shanghai University, Shanghai, China

G. Li
Shanghai Marine Equipment Research Institute, Shanghai, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_13

1 Introduction

Journal bearings owing to its specific advantages are more and more widely used in high-speed rotating machinery. Dynamic characteristics which is the most important embodiment of bearing always have significant effects on the stability of rotor-bearing system. So its research work has been concerned by the academic and engineering circles.

In many cases, Dynamic coefficients are assumed on the linear simplification. In recent years, many methods have been proposed to change the dynamic characteristics of the bearings. Surface texture is a specific method that can be processed by laser. LST (laser surface texturing) provides a relatively cheap and rapid method to create textured surface. Laser which is a fast and environmentally-friendly method provides excellent control of the shape and size of the dimples that allows realization of optimum parameters [1].

In the past few years, a large number of studies have been conducted that considered the influence of surface texture on the performance of bearing systems [2–4]. Surface texture has been widely applied in automotive industry to minimize friction and leakage in the seal industry and to reduce friction of piston rings. It can be applied to improve the external load carrying capacity in thrust bearings and to enhance the performance of journal bearings. The results reporting that surface texture is power in friction reduction and enhancement of load-carrying capacity significantly [5–7].

There are a large number of theoretical studies of surface texture. The positive influence of artificial surface texturing on wear resistance, enhancing the load capacity and friction coefficient of mechanical seals [8]. Etsion and Burstein [9] researched the effect of shallow pores on the operational behavior of the mechanical face seals. They solved the Reynolds equation numerically using half-Sommerfeld boundary condition to predict the performance of mechanical seals comprising hemispherical pores.

Theoretical and experimental researchs of a laser textured thrust bearing had been done by Marian et al. [10, 11] and a good agreement was found between the theoretical model and the experimental data for the fluid film thickness and friction torque.

There are also many experimental studies on surface texture have been reported. Experimental studies have shown that textured bearing surfaces exhibit a higher load carrying capacity [12]. A comparison of partially and fully textured surfaces comprising micro-dimples in parallel thrust bearings has been studied by Brizmer [12] through employing a numerical approach. A large number of numerical simulations results reveal that the area density of pores, the textured portion, the pores height ratio and the bearing length to width ratio are the most affecting parameters in computing bearing load capacity.

There is no experimental study has been reported about the effect of surface texture on the performance of dynamic performance of journal bearing.

Therefore, the purpose of the present paper is to investigate experimentally the validity of the model described in Khatri [13] by testing practical journal bearings and comparing the performance of LST bearings with that of the theoretical predictions and with the performance of standard non-textured bearings. Experimental results have been compared with a theoretical model. The experimental results are in good agreement with the theoretical model under relevant operating conditions.

2 Experimental Method

2.1 Experiment Apparatus

The same bearing geometrical parameters and test rig studied by Jiexi Shen [14] are used here for the experiment. Figure 1 shows the 3D model of the journal bearing test rig. The test rig is mainly composed of the following parts:

1. Two pedestals to support the test rotor through two same ball bearings that are lubricated through the oil supply system.
2. An electromotor used to drive the rotor. The maximum speed is 3000 rpm.
3. A pneumatic loader to apply a static load to the stator in the positive y direction. The maximum available load is 5 kN.
4. Dynamic load system, by 90° two degrees of the vibration exciter, the use of non-contact exciter, the machine of the vibration exciter is JZQ-70, and the maximum output exciting force is 700 N.
5. Data acquisition system. Two accelerometers were used to test the vibration amplitude of bearing housing, and four eddy displacement sensors were used to test displacement of the journal bearing in the X and Y directions respectively.

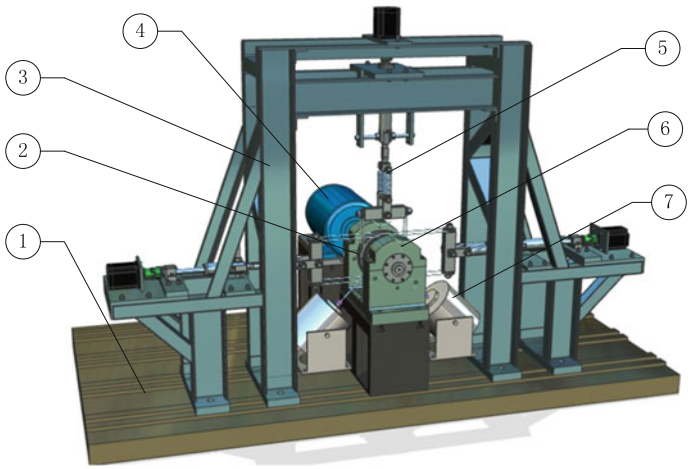


Fig. 1 Overview of the test rig. **1** foundation; **2** test bearing; **3** test frame; **4** drive system; **5** static load equipment; **6** rotor and support; **7** excitiers

6. Oil supplying system. ISO VG32 turbine oil is delivered to the journal bearing at the temperature of 30°. The flow rate of supply oil controlled through the driven valves and heat exchanger system.

The aim of the experiment are (1) test the effect of texture on the stability of bearing, (2) demonstrate the bearing dynamic performance of the textured bearing and (3) test the effect of texture on the vibration of rotor-bearing system.

2.2 Test Bearing

In order to study the effect of texture on the dynamic characteristics of the bearing, the experiments on the bearing with non-textured and textured are carried out. Etsion [9] through the study of the hemispherical micro pits on the mechanical seal face fluid between the effect of dynamic pressure bearing capacity of the micro pits depth and diameter ratio of fluid dynamic pressure effect is very significant, the optimal aspect ratio of less than 0.05, while the influence of area density is small, the area density is better about 0.2. Talalghil [15–17] calculated the effect of texture dimensions on the bearing performance study, also studied the effects of different texture distribution pattern on the lubrication performance, found all the texture lubrication performance becomes worse, and if the oil film pressure drop at the part of the design to improve the performance of the bearings of the texture.

The parameters of the experimental bearings are shown in Table 1. In this paper, we select the surface texture of the square concave type as the radial sliding bearing. Texture parameters: the rectangular pit size of 500 × 500 μm, the depth of 20 μm, the area ratio of 25% and the distribution of 0°–160° in lower bearing, the distribution of texture in the bearing surface as shown in Fig. 2.

2.3 Measuring Method

Four Eddy current displacement sensors are used in this test to record the rotor displacement of the x and y directions. The install location of the eddy current displacement sensors is shown in Fig. 3. The sampling frequency of the sensor is 2560/s. Four eddy current displacement sensors measure the displacement of the bearings vertical and horizontal direction respectively, 1 and 3 are the absolute

Table 1 Parameter of rotor-bearing system

Bearing diameter (mm)	Bearing length (mm)	Bearing clearance (mm)	Preload factor	Type of lubricant	Oil inlet temperature °C
70	49	0.08	0.1	ISO-VG32	30

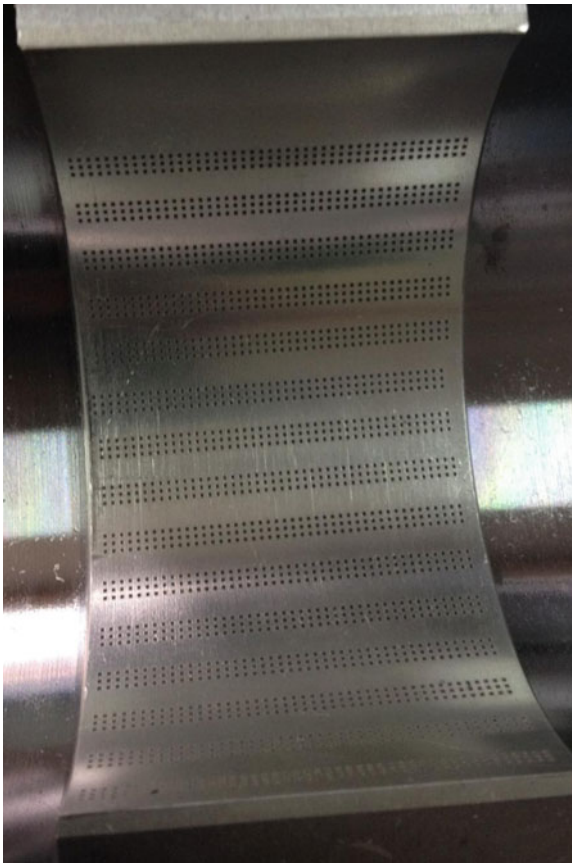


Fig. 2 A photograph of textured bearing

displacement of rotor, 2 and 4 are relative displacement of the bearing and foundation. The sensors are arranged at one end of the motor. Two eddy current transducers are needed to be installed perpendicular and coplanar with each other.

Accelerometers (PCB Model: 352C03; Measurement Range: 500 lb (2.224 kN)) produced by PCB company are also used in this test. The sensors are arranged on one end of the motor. The measured data of the bearings will be eventually obtained by signal sampling and processing system from Bruel & Kjaer PULSE.

2.4 Dynamic Characteristics of Journal Bearing

Stiffness, damping, and added-mass coefficients are very important in modeling fluid-structure interaction forces for rotor dynamic analysis. These coefficients are typically expressed in the following linearized force-displacement bearing model:

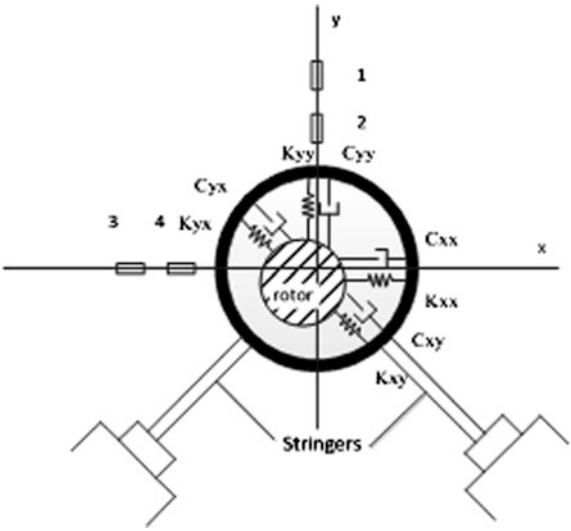


Fig. 3 Location of eddy current displacement sensors. 1, 2, 3, 4 are eddy current displacement sensors

$$-\begin{bmatrix} f_{bx} \\ f_{by} \end{bmatrix} = \begin{bmatrix} k_{xx} & k_{xy} \\ k_{yx} & k_{yy} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} + \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix} \begin{bmatrix} \Delta \dot{x} \\ \Delta \dot{y} \end{bmatrix} \tag{1}$$

Childs and Hale [18] explained how the rotor dynamic coefficients are estimated from the measurements. In this paper, we use the same method to obtain the dynamic coefficients.

Mechanical model with stiffness [K], damping [C], and added-mass [M] as

$$-\begin{bmatrix} f_x - M_s\ddot{x}_s \\ f_y - M_s\ddot{y}_s \end{bmatrix} = \begin{bmatrix} k_{xx} & k_{xy} \\ k_{yx} & k_{yy} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} + \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix} \begin{bmatrix} \Delta \dot{x} \\ \Delta \dot{y} \end{bmatrix} - \begin{bmatrix} M_{xx} & M_{xy} \\ M_{yx} & M_{yy} \end{bmatrix} \begin{bmatrix} \Delta \ddot{x} \\ \Delta \ddot{y} \end{bmatrix} \tag{2}$$

Equation (2) applying a fast Fourier transform (FFT) produce.

$$\begin{bmatrix} F_x - M_s A_x \\ F_y - M_s A_y \end{bmatrix} = -\begin{bmatrix} H_{xx} & H_{xy} \\ H_{yx} & H_{yy} \end{bmatrix} \begin{bmatrix} D_x \\ D_y \end{bmatrix} \tag{3}$$

Independent excitation in orthogonal directions x and y directions produces the following four independent equations:

$$-\begin{bmatrix} F_{xx} - M_s A_{xx} & F_{xy} - M_s A_{xy} \\ F_{yx} - M_s A_{yx} & F_{yy} - M_s A_{yy} \end{bmatrix} = -\begin{bmatrix} H_{xx} & H_{xy} \\ H_{yx} & H_{yy} \end{bmatrix} \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix} \tag{4}$$

Relationships between the frequency response function H_{ij} and the rotor dynamic coefficients are as follows:

$$H_{ij} = K_{ij} - \Omega^2 M_{ij} + j(\Omega C_{ij}) \tag{5}$$

3 Experimental Results and Discussion

3.1 Dynamic Characteristics of Journal Bearing and Textured Bearing Under Different Loads

In order to evaluate the dynamic characteristics of the partially textured bearing compared to a non-textured journal bearing, the stiffness coefficients and damper coefficients were calculated for both bearings. Figure 4 presents the experimental dynamic stiffness coefficients under 1000–2500 N and at 1200 rpm. Figure 4 clearly demonstrates that the partly textured journal bearing significantly increases the stiffness coefficients of bearing vis a vis non-textured journal bearing. The increase in the value of stiffness is consistent with observation of earlier reported study of Khatri and Sharma [13]. From Fig. 4a, it can be observed that K_{xx} increases slightly with an increase of the applied speed in the low load condition, but reduce with an increase of the applied speed in the heavy load condition. When the load is 2000 N, K_{xx} is the biggest. At the given load range, the value of K_{xx} is enhanced 60.05–340.66% by using bearing with the partly textured surface. K_{yy} of the textured bearing is similar to K_{xx}. K_{yy} of textured bearing increases with an increase of the applied speed in the low load condition. At the given load range, the value of K_{yy} is enhanced by 235.92–840.78% by using bearing with partly texture surface.

Cross-coupled stiffness of the bearing can affect the stability of the bearing. From Fig. 4b, it can be observed that the cross-coupled stiffness of the bearing can affect the stability of the bearing. Figure 4b clearly demonstrates that the relation between cross-coupled stiffness and load is not monotonic, but the cross-coupled stiffness coefficients of the non-textured bearing are larger than that of the surface

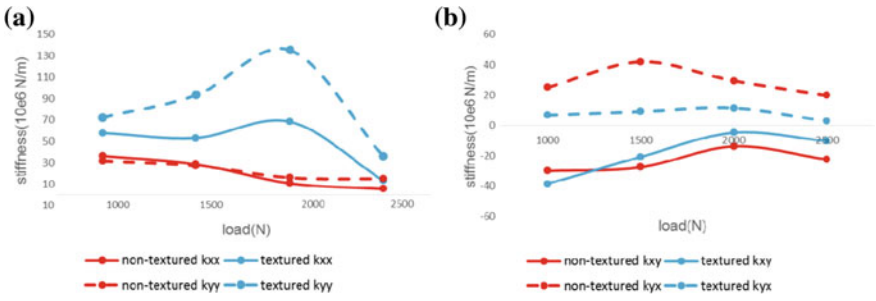


Fig. 4 Stiffness at 1000–2500 N and 1200 rpm. **a** is direct stiffness, **b** is cross-coupled stiffness

textured bearing. According to the author described that [19] the destabilizing effect of cross coupling is directly proportional to the area of the whirl orbit, which can be written as follows:

$$E_{cyc} = A(K_{xy} - K_{yx}) \tag{6}$$

In Eq. [6], A represents the area enclosed by the elliptical orbit while parameters K_{xy} and K_{yx} mean cross-coupled stiffness and E_{cyc} represents energy per cycle imparted to the shaft [20]. Therefore, in some way, the surface texture increases the stability of the rotor through affects the cross stiffness of the bearing.

Figure 5 presents the experimental dynamic damping coefficients at 1000–2500 N and at 1200 rpm. The damping coefficient is related to the ability of a journal bearing to absorb shaft vibration. From Fig. 5a, it can be observed that, by using surface texturing, the value of damping coefficient improves significantly as compared to the non-textured bearing.

The improvement of the journal bearing characteristics consists in increasing the minimum film thickness, improving the lubricant film presence region.

It can easily be seen that C_{xx} and C_{yy}.reduce slightly with an increase of the applied speed in the low load condition. At the given load range, the damping coefficients of the surface textured bearing are larger than that of the non-textured bearing. The value of C_{xx} and is enhanced by 419.37–1811.14%. As a whole, it can be seen that the damping coefficients of the surface textured bearing are larger than that of the non-textured bearing.

Texturing the inlet region of the contact located at the increasing part of the pressure field improves the pressure field distribution, enhances the carrying capacity and increases the minimal film thickness The damping coefficient of the bearing is increased.

The cross-coupled damping of the bearing can affect the stability of the bearing. Figure 5b clearly demonstrates that the relation between cross-coupled damping and loading is not monotonic, but the damping coefficient of the surface textured bearing is larger than that of the non-textured bearing.

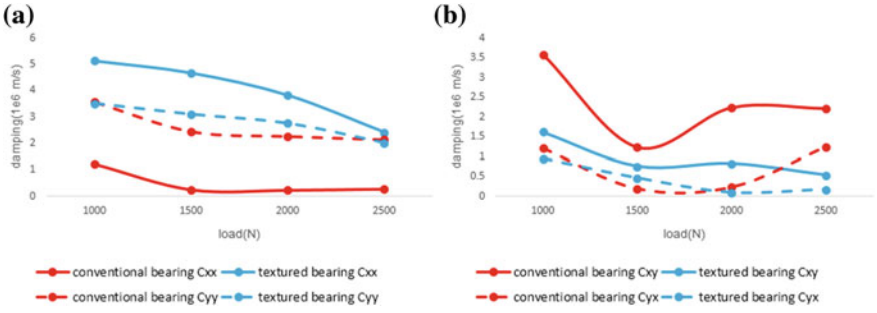


Fig. 5 Damping at 1000–2500 N and 1200 rpm. **a** is direct damping, **b** is cross-coupled damping

Obviously, when the micro texture area was arranged in the inlet region of the contact located at the increasing part of the pressure field, stiffness coefficient and damping coefficient of the bearing is larger than that of the non-textured bearing.

By comparing the experimental results with the values of textured and non-textured, we get the conclusion that the surface texture can improve the stiffness coefficient and damping coefficient of the journal bearing.

3.2 *Vibration Characteristics of Journal Bearing and Textured Bearing Under Different Loads*

The bearing housing vibration level is important to evaluate rotating machinery quality. In order to study the effect of surface texture on the vibration of bearing housing, Acceleration sensors are placed in horizontal and vertical directions on the bearing housing. It is worth mentioning here that the value of acceleration also includes the vibrations generated due to the rotation of the disk and the drive system. Figure 6 presents the experimental vertical acceleration at load from 1000 to 2500 N.

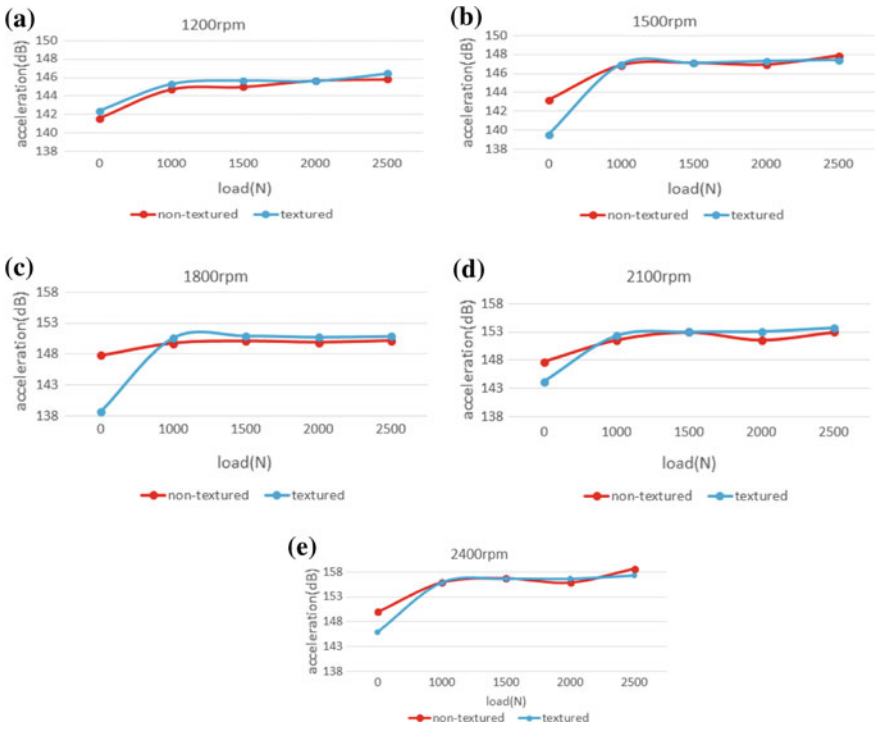


Fig. 6 Acceleration amplitude of bearing housing. **a** 1200 rpm, **b** 1500 rpm, **c** 1800 rpm, **d** 2100 rpm and **e** 2400 rpm

From Fig. 6, it can be clearly seen that the value of vibration amplitude of textured journal bearing is lower than the non-textured bearing in high speed and light load circumstance. The surface texture can decrease the value of vibration amplitude 9 dB in 1800 rpm and no load circumstance. The different value of vibration between non-textured bearing and textured bearing becomes smaller as the load increases, and the difference becomes not obvious after the load is 1000 N. In the case of light load, the surface textured bearing has a certain effect on the damping. In the case of heavy load, the surface texture has little effect on the damping.

Texturing the contact inlet region has a positive effect on the film thickness at light load. Surface microcavity can hold more lubricant oil. The damping coefficients of the textured journal bearing are larger than the damping coefficients of the non-textured bearing, the damping coefficients are positively correlated with the vibration. Surface texture increases the damper coefficients of bearing and decreases bearing housing vibration. Surface texture has an influence on bearing housing vibration at heavy load.

By the analysis above, the experimental results prove that surface texture can be utilized in the field of vibration control for a rotating machine. So we get the conclusion that the reasonable design of the surface texture can achieve the effect of shock absorption.

4 Conclusions

This article is mainly for the experimental measurement of dynamic characteristics of the surface textured journal bearing, and the dynamic behavior is compared between surface textured journal bearing and non-textured journal bearing. Some conclusions can be obtained as follows:

(1) The textured journal bearing provides a higher value of fluid film stiffness coefficients and damping coefficients.
(2) The textured journal bearing can reduce the vibration of rotor-bearing system. In the case of light load, the experimental study shows that using surface texture is a feasible and effective strategy to suppress vibration.
(3) This experimental result is useful for the determination of stiffness coefficients and damping coefficients of bearing the design of such bearings. The experimental results show that the appropriate arrangement of the textured area on the contact inlet region can change the dynamic characteristics of bearing.

Acknowledgements The authors acknowledge the support from the Shanghai Science and Technology Committee (no. 14111103703) and Shanghai Key Laboratory of Intelligent Manufacturing and Robotics.

Appendix

M_S	The stator mass
$\ddot{x}_s$, $\ddot{y}_s$	The stator acceleration in X and Y direction
f_x, f_y	The force components induced by the electromagnetic vibration exciter
Δx, Δy	The bearing displacement in X and Y direction
$\Delta\dot{x}$, $\Delta\dot{y}$	The bearing velocity in X and Y direction
$\Delta\ddot{x}$, $\Delta\ddot{y}$	The bearing acceleration in X and Y direction
D_i	The fourier transform of the measured displacements in the i direction
H_{ij}	The direct and cross-coupled dynamic-stiffness coefficients for the bearing
F_i	The complex fourier transformed values for the measured force
A_i	The complex fourier transformed values for the acceleration components

References

1. Brizmer V, Kligerman Y, Etsion I (2003) A laser surface textured parallel thrust bearing. Tribol Trans 46:397–403
2. Etsion I (2004) Improving tribological performance of mechanical components by laser surface texturing. Tribol Lett 17:733–737
3. Scaraggi M, Mezzapesa F, Carbone G, Ancona A, Tricarico L (2013) Friction properties of lubricated laser-microtextured-surfaces:an experimental study from boundary-to hydrodynamic tolubrication. Tribol Lett 49:117–125
4. Qiu M, Delic A, Raeymaekers B (2012) The effect of texture shape on the load-carrying capacity of gas-lubricated parallel slider bearings. Tribol Lett 48:315–327
5. Lu X, Khonsari MM (2007) An experimental investigation of dimple effect on the stribeck curve of journal bearings. Tribol Lett 27:169–176
6. Rao T, Rani A, Nagarajan T, Hashim F (2014) Analysis of couple stress fluid lubricated partially textured slip slider and journal bearing using narrow groove theory. Tribol Int 69:1–9
7. Etsion I, Halperin G, Brizmer V, Kligerman Y (2004) Experimental investigation of laser surface textured parallel thrust bearings. Tribol Lett 17:295–300
8. Etsion I, Kligerman Y, Halperin G (1999) Analytical and experimental investigation of laser-textured mechanical seal faces. Tribol Trans 42(3):511–516
9. Etsion I, Burstein L (1996) A model for mechanical seals with regular microsurface structure. Tribol Trans 39(3):677–683
10. Marian VG, Kilian M, Scholz W (2007) Theoretical and experimental analysis of a partially textured thrust bearing with square dimples. Proc Inst Mech Eng, Part J: J Eng Tribol 221 (7):771–778
11. Marian VG, Gabriel D, Knoll G et al (2011) Theoretical and experimental analysis of a laser textured thrust bearing. Tribol Lett 44(3):335–343
12. Brizmer V, Kligerman Y, Etsion I (2003) A laser surface textured parallel thrust bearing. Tribol Trans 46(3):397–403
13. Khatri Chandra B, Sharma Satish C (2016) Influence of textured surface on the performance of nonrecessed hybrid journal bearing operating with non-Newtonianlubricant. Tribol Int 95:221–235
14. Jiexi S, Xin X (2016) Experimental analysis of dynamic oil film pressure of tilting-pad journal bearings Triblo

15. Talaighil N, Maspeyrot P, Fillon M, Bounif A (2007) Effects of surface texture on journal bearing characteristics under steady state operating conditions. Proc Inst Mech Eng, Part J J Eng Tribol 221(6):623–633
16. Talaighil N, Fillon M, Maspeyrot P (2011) Effect of textured area on the performances of a hydrodynamic journal bearing [J]. Tribol Int 44:211–219
17. Talaighil N, Maspeyrot P, Fillon M, et al (2007) Hydrodynamic effects of texture geometries on journal bearing surfaces. In: The 10th international conference on Tribology, pp 47–52
18. Al-Ghasem A, Childs D (2006) Dynamic coefficients; measurements versus predictions for a high speed flexure-pivot tilting-pad bearing (Load-Between-Pad Configuration). ASME J Eng Gas Turbines Power 128(4):896–906
19. Zeidan FY, Herbage BS (1991) Fluid film bearing fundamentals and failure analysis. In: Proceedings of the 20th turbo machinery symposium, pp 161–186
20. Maurice L, Adams JR (2001) Rotating machinery—from analysis to troubleshooting. Marcel Dekker, New York

The Effect of Temperature on Mechanical Properties of Polypropylene

Guiqin Li, Junjie Li, Jun Wang, Jiemin Feng, Qing Guo, Junlong Zhou and Peter Mitrouchev

Abstract The significant diversity exists among the mechanical properties of PP polymers in range from high temperature to low temperature. With the increasing application of PP polymers in the automobile industry, the mechanical properties of PP polymers which depending on temperature are extensively studied by a growing number of scholars. The presented study concentrates on determining the mechanical properties depending on temperature by the use of mathematical method. With the focus on automotive applications, the elasticity constants values have been determined in the range from −30 up to 110 °C using uniaxial tensile. Experimental results indicate that the elasticity present a dramatic decrease with the increase in temperature, Based on the experimental investigation and the theoretical research by domestic and foreign scholars. The properties like the elasticity are mathematically described depending on the temperature in different mathematical methods and the influence of temperature on elastic modulus was summarized.

Keywords PP polymer · Temperature effect · Elastic modulus

1 Introduction

Polypropylene (PP) reinforced thermoplastic materials show high load structure applications which make it widely used in the automotive industry [1], the mechanical properties of polymers exist significant diversity in range from high

G. Li (✉) · J. Li · J. Wang · J. Feng
Shanghai Key Laboratory of Intelligent Manufacturing and Robotics,
Shanghai University, Shanghai 200072, China
e-mail: leeching@shu.edu.cn

Q. Guo · J. Zhou
Shanghai Yanfeng Jingqiao automotive trim systems Co. Ltd, Shanghai, China

P. Mitrouchev (✉)
University Grenoble Alpes, G-SCOP, 38031 Grenoble, France
e-mail: peter.mitrouchev@g-scop.inpg.fr

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_14

temperature to low temperature, It is strongly necessary to study the physical properties of such materials at different temperatures [2]. PP polymers are in a glassy state at low temperatures or are elastomeric, while at higher temperatures are rubber or entropy, and the transition temperature range between glass and rubber is defined as the glass transition zone, glass the conversion temperature is T_g [3, 4]. When the temperature is low, the polymer glassy state is characterized by elastic deformation when subjected to external force. When the temperature is high, the polymer material appears in the rubber state can occur within the intramolecular rotation and the chain and the branch of the displacement that Brownian movement, then macromolecules once subjected to external force can take the extended state. If the elongation of the molecular chain deviates from the best curl state, then the system can increase, while the entropy is reduced. On the other hand, due to the inherent vibration of the material inside the material, the elastic retraction force of the molecule itself tends to tend to return to the crimp state slowly. The polymer material in this state is characterized by rubber.

In order to provide reliable mechanical properties with mathematical descriptions of temperature changes, Ha and Springer proposed a nonlinear relationship between elastic properties and temperature, Gibson et al. demonstrated a mechanical properties that included changes in the glass transition region. The hyperbolic relationship [4, 5] applies to all temperature segments and establishes an empirical formula model, in addition to Hufenbach et al. also proposed a multi-linear functional relationship. Based on these studies, the relationship between the performance parameters and the temperature of the PP polymer was obtained by tensile test.

2 Experimental Investigation

2.1 Material

The material used in the present study was the addition of 20% talc of polypropylene and EPDM rubber polycarbonate (PP + EPDM-T20) which are two grades (TRC352NC3/TRC787 N). Specimens were machined for the ISO527-1A direct injection of the sample, the test length and width were 50 and 12.7 mm. The length of each sample was measured with a caliper and kept at test temperature for more than 20 min before testing, see Fig. 1.

2.2 Tensile Testing

The tensile test was carried out using a Zwick/Roell Z020 tensile tester, see Fig. 2. In order to adjust the ambient temperature, a climatic chamber equipped with a high temperature heating furnace and the low temperature liquid nitrogen was used. The

Fig. 1 The test samples

Fig. 2 The tensile tester

use of multiple thermocouple elements in the climate chamber and reference samples for on-line temperature measurement to ensure that the sample is tested at the specified temperature. The uniaxial tensile test was carried out at a temperature of −30, −10, 23, 60, 80 and 110 °C to obtain a stress of uniaxial tension Strain curve and the maximum tensile stress, the maximum tensile stress under the strain of two parameters, the processing of elastic modulus.

3 Results and Discussion

The elastic modulus data are obtained in range from 0.05 up to 0.25% of the strain by the stress–strain curves which are plotted by the uniaxial tensile test. The values of the elastic modulus are shown in Table 1.

An empirical relationship with a hyperbolic function for modelling the changing of the mechanical property values within the transition regions is used by Gibson, it is simple expressed as (1),

$$E(\mathrm{T}) = a * \tanh\left(\frac{T+b}{c}\right) + d \tag{1}$$

Where the coefficients a, b, c, d are independent of temperature and are uniquely determined by the nature of the material.

In addition, a multi-linear approach which seems also reasonable with respect to the changing material properties in different temperature is proposed here using a more concise formula section by section:

$$P(\mathrm{T}) = a_p T + b_p \tag{2}$$

Table 1 The experimental values of modulus (Unit: MPa)

Grade	−30 °C	−10 °C	23 °C	60 °C	80 °C	110 °C
352NC-1	3380.07	2210.01	1683.24	300.48	286.79	142.10
352NC-2	3251.57	2390.02	1693.52	483.58	259.79	165.83
352NC-3	2910.34	2298.12`	1672.61	465.69	226.63	149.90
787N-1	2588.44	2433.63	1232.10	379.80	147.81	103.47
787N-2	2939.99	2600.29	1298.34	372.32	194.04	131.42
787N-3	3053.95	2485.16	1262.76	454.39	142.40	111.28

Table 2 Adj. R-square values

Modelling approaches	352NC	787N
Hyperbolic approach	0.97299	0.99158
Ha/springer approach	0.95429	0.96357
Multi-linear approach	0.90468	0.86468

Ha and Springer [6] proposed a power function of the relationship model, the expression is shown as (3):

$$E(\mathrm{T}) = T_0 \tag{3}$$

The data were solved simultaneously in least squares by fitting the model to the experimental. The comparison of the measured values with the modelling is presented in Fig. 1. The fitting correlation coefficients (Adj. R-Square) are shown in Table 2 which means the good agreement of the curve fitting is adjacent to 1.

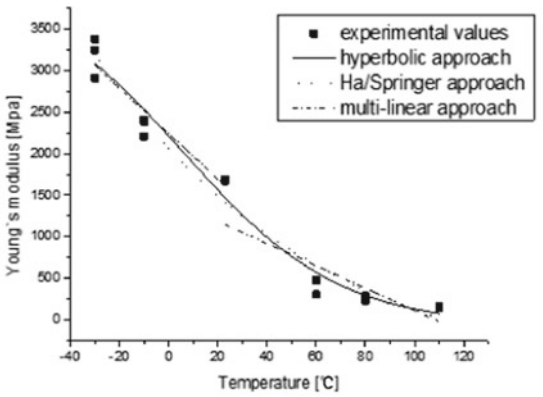
The data were solved simultaneously in least squares by fitting the model to the experimental. The comparison of the measured values with the modelling is presented in Fig. 3. The fitting correlation coefficients (Adj. R-Square) are shown in Table 2 which means the good agreement of the curve fitting is adjacent to 1.

In general, the presented multi-linear and hyperbolic approach represents the temperature dependence more accurate. Ha-Springer approach is not suitable for the whole temperature range. Compared with the Adj. R-Square of the modelling approach, the fitting effect of hyperbolic tangent function is the best.

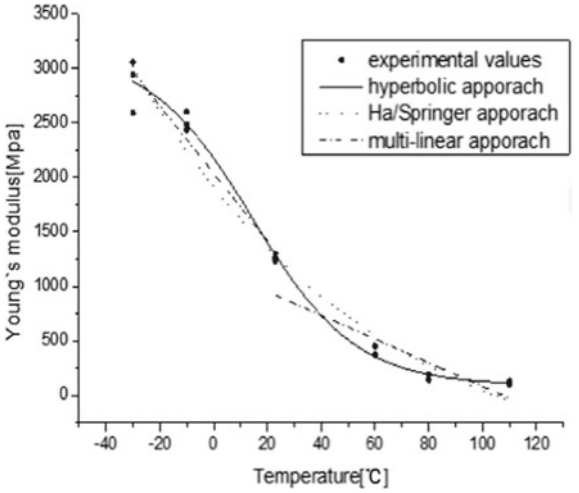
After fitting with OrginPro, query its fitting correlation coefficient Adj. R-Square as shown in Table 2. The closer the value is to 1, the higher the degree of fit. After fitting with OrginPro, query its fitting correlation coefficient Adj. R-Square as shown in Table 1. The closer the value is to 1, the higher the degree of fit. After fitting with OrginPro, query its fitting correlation coefficient Adj. R-Square as shown in Table 1. The closer the value is to 1, the higher the degree of fit.

4 Conclusion

The mechanical properties of polymers were tested in the temperature range from −30 up to 110 °C. Based on the theoretical research by domestic and foreign scholars, the elastic properties of polypropylene is summarized and described as a function of temperature. Compared with the three modelling, the fitting effect of hyperbolic tangent function is the best. The model has a good degree of fitting and more intuitive. This rule can reduce the difficulty of the test, shorten the test cycle, for the automotive industry and other sectors of the use of thermoplastic materials to provide a more convenient basis for performance prediction.



(a) 352NC



(b) 787N

Fig. 3 Comparison of measured values with the modelling approaches

Acknowledgements This research is partly supported by Shanghai Automotive Industry Science and Technology Development Fund (NO. 1512).

References

1. Yang T (2013) Present status and prospects of plastic materials used in automobile industry. New Chem Mater 41(5):25–29
2. Li G, Chen y (2016) A constitutive model of polymer at different temperature. J Mech Eng 52 (12):78–81

3. Gibson AG, Otheguy-Torres ME (2010) High temperature and fire behavior of continuous glass polypropylene laminates. Composites 41(12):19–31
4. Odegard G, Kumosa M (2000) Elastic-plastic and failure properties of a unidirectional carbon/PMR-15 composite at room and elevated temperatures. Compos Sci Technol 29:79–88
5. Mouritz AP, Gibosn AG (2006) Fire properties of polymer composite materials. Solid Mech Appl 1:43
6. Ha Springer (2009) Determination of mechanical properties of glass-epoxy composites in high temperature. Polym Compos 30(10):37–41

The Effect of Extensometers on the Mechanical Properties of the Polypropylene Under Uniaxial Tensile Loading

Guiqin Li, Hao Chen, Jun Wang, Jiemin Feng, Qing Guo, Junlong Zhou and Peter Mitrouchev

Abstract The elastic modulus and yield strain of polypropylene were obtained by the tensile test at different temperature. Compare the physical performance parameters which were obtained with or without using the extensimeter, a linear relationship between the two schemes is presented in this article. The coefficient of the linear relationship was calculated by the two sets of data which are obtained with or without using the extensimeter, and the presented linear fitting represents the results more accurate. It greatly reduces the difficulty and intensity of the experiment and provides a more convenient method for the performance testing and testing of high polymer in automobile manufacturing industry.

Keywords Polymer · Extensometer · Tensile test · Physical performance

1 Introduction

With the increasing use of high polymers in the automotive industry, the physical performance testing becomes particularly important. This paper concentrates on studying the influence of the elastic modulus E and strain ε of high polymer with extensometers in tensile test. Non-contact video test principle is completely different form the contact strain extensometer access method. Non-contact principle is

G. Li (✉) · H. Chen · J. Wang · J. Feng
Shanghai Key Laboratory of Intelligent Manufacturing and Robotics, Shanghai University, Shanghai 200072, China
e-mail: leeching@shu.edu.cn

Q. Guo · J. Zhou
Shanghai Yanfeng Jingqiao Automotive Trim Systems Co. Ltd, Shanghai, China

P. Mitrouchev (✉)
University Grenoble Alpes, G-SCOP, 38031 Grenoble, France
e-mail: peter.mitrouchev@g-scop.inpg.fr

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_15

using the method of sub-pixel extensometer in contactless way to measure the deformation of specimen and its advantages are non-contact, no wear, without introducing additional error in measurement, high measuring accuracy, and it can measure the tensile deformation process, which can be measured materials of various limit parameters.

At present, the physical performance parameters can be obtained accurately by the extensometer in the uniaxial tensile test based on the long-term experimental studies by more and more domestic and international scholars [1, 2]. Such as Noel [3], they have validated the mechanical properties of the polymer at different temperatures. The performance and morphology in the tensile test of high polymer have analyzed by Lovinger [4]. By using the application of no-contact displacement in the stretching test, Zhang [5] have obtained the physical performance parameters which is similar to the expected results. The purpose of this paper is to provide a formula to describe the transformational relationship under the conditions with or without the extensometers in the tensile test. The real mechanical properties can be calculated by the linear formula in the absence of extensometers which can reduce the difficulty of uniaxial tensile test.

2 Test Method for Properties of Plastic Materials at High and Low Temperature

The high and low temperature tensile test was carried out on a Zwick/Roll Z020 testing machine which was loaded with displacement control. The values of collect load, displacement, strain and time was measured at −30, −10, 23, 60, 80, 110 °C. Specific materials as follow: (1) Add 20% talc of polypropylene material (PP + T20); (2) Addition of 20% talc polypropylene and EPDM (PP + EPDM + T20); (3) Polybutylene terephthalate (PBT).

Specimens were machined for the ISO527-1A direct injection of the A4 plate sample. The total length of dumbbell-like specimen is 100 mm, the test gauge length is 50 mm and the width is 12.7 mm. The uniaxial tensile test was testing by the cross speed of 1 mm/min in a temperature range from −30 to 110°C. The test temperature was gradually increased until the appoint temperature. The Specimen was preheated at least 30 min in the required heat chamber and kept 10 min on the test machine before testing loading. These serve to ensure that the preheat treatment will not affect the mechanical properties. Then the tensile test scheme (sampling frequency 20 frames/sec) of the video extensometer was used to compare the experimental data directly with the contact strain acquisition method to find out the relationship.

3 Results and Discussion

The elastic modulus data are obtained in range from 0.05 up to 0.25% of the strain by the stress-strain curve which is plotted by the uniaxial tensile test. As can be seen form the Fig. 1, the elastic modulus presents a dramatic decrease with the increase in temperature. The values of the elastic modulus obtained by using the extensometer are different from the values without using the extensometer.

It is obviously observed that it exist a particular ratio between the two sets of data at the low temperature. It can be assumed that the ratio is equal to the values obtained by the extensometer divided by the values without the extensometer. The ratios of the elastic modulus are shown in Fig. 2.

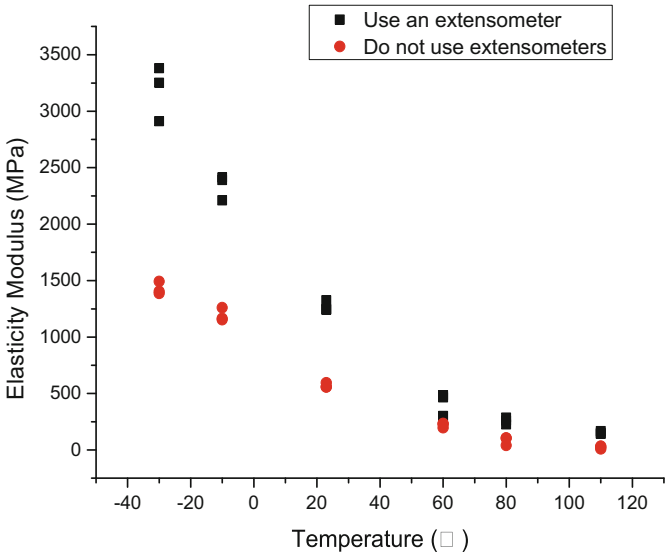
It is apparent that for several different types of polymeric materials, the change of E at different temperatures is different. In the range of −30–60 °C, the ratio of using extensometer or not to E is close to the linear relationship, and the scale factor is close to 2.0 ± 0.15. And the ratio of the extensometer to the E at 60–110 °C is an exponential relationship. During the uniaxial tensile test at different temperatures, the displacement of the chuck on both sides of the specimen is large. Therefore, the strain in the test without extension is larger than that using the extensometer. The E is less than that of using an extensometer.

4 Verification Test

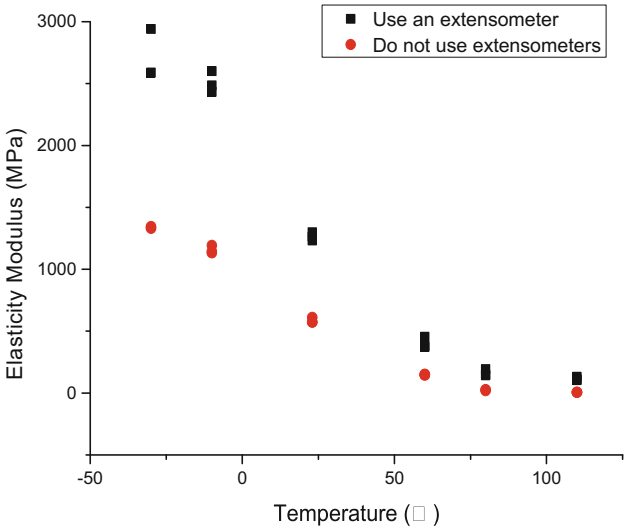
In order to verify the relationship between the mechanical performance parameters under the uniaxial tensile test of the extensometer and the uniaxial tensile test without the use of the extensometer, the test material is a thermoplastic of another reinforced polypropylene. The elastic modulus and strain at 6 kinds of temperatures were obtained. Tables 1 and 2 list the elastic modulus and strain parameters for the use of an extensometers at six kinds of temperature. Table 3 shows the ratio of E and ε using the extensometer and without the extensometer. It can be seen from the table, using of extensometer or not the uniaxial tensile test at different temperatures is a certain relationship.

5 Conclusion

The values of the E obtained by using the extensometer are different from the values without using the extensometer. The relationship of the E between the two conditions form −30 to 60 °C is close to the linear interpolation, and the ratio of the E from 60 to 110 °C is an exponential relationship $k_1 = 0.397e^{0.0272x}$. The ratio of the ε between the two conditions by the same material is linear. The real ε can be calculated by multiplying the scaling factor under the condition of testing without

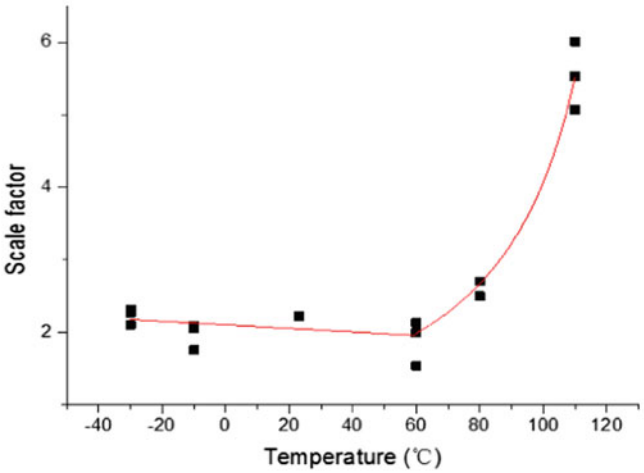


(a) PP+EPDM+T20 (Basell TRC352N C3)

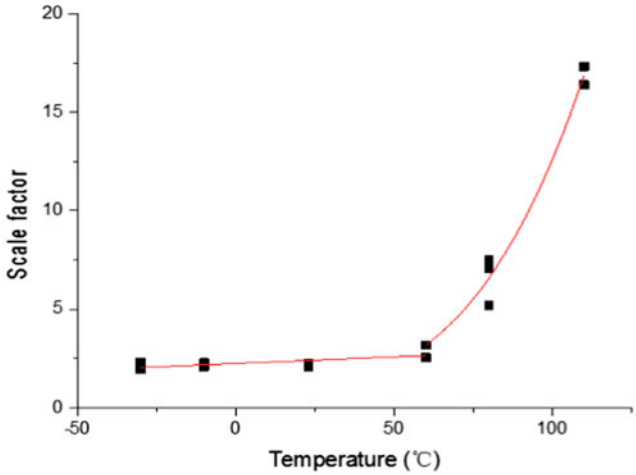


(b) PP+EPDM+T20 (Basell TRC352N C3)

Fig. 1 The elastic modulus for different temperature



(a) PP+EPDM+T20 (Basell TRC352N C3)



(b) PP+EPDM+T20 (BasellTRC352NC3)

Fig. 2 The ratios of the E with or without the extensometer at different temperature

Table 1 Properties by using extensometer

Temperature/°C	E/MPa	ε
−30	3097.70	0.0232
−10	2506.14	0.0235
23	1802.49	0.0493
60	320.60	0.0647
80	235.88	0.0672
110	134.68	0.0748

Table 2 Properties without extensometer

Temperature/°C	E/MPa	ε
−30	1369.30	0.0469
−10	1178.70	0.0458
23	606.21	0.1006
60	182.50	0.1217
80	86.31	0.135
110	15.40	0.1526

Table 3 The ratios of E and ε with or without the extensometer at different temperature

Parameter	−30 °C	−10 °C	23 °C	60 °C	80 °C	110 °C
k_1	2.26	2.126	2.373	2.357	2.733	8.744
k_2	0.494	0.514	0.491	0.532	0.496	0.49

k_1 present the ratio of the elastic modulus, k_2 present the ratio of the yield strain

the extensometer. It greatly reduces the difficulty and intensity of the experiment and provides a more convenient method for the performance testing of high polymer in automobile manufacturing industry.

Acknowledgements This research is partly supported by Shanghai Automotive Industry Science and Technology Development Fund (NO.1512).

References

1. Zhan-fu WANG, Li-ping XIE (2011) Design of non-contact video strain measurement system. Tool Technol 45(8):91–94
2. Tong S, Li S, Yao C (2015) Method for controlling terminal device, and wearable electronic device: WO/2015/081558
3. Noel III † OF, Carley JF (1975) Properties of polypropylene-polyethylene blends. Polym Eng Sci 15(2):117–126
4. Lovinger AJ, Williams ML (1980) Tensile properties and morphology of blends of polyethylene and polypropylene. J Appl Polym Sci 25(8):1703–1713
5. Zhang W-q, Bai H (2011) Comparison of tensile test results of non-contact displacement/strain video measuring instrument, (2011) Chinese Mechanical Congress 2011 and the 100th anniversary of the birth of Qian Xuesen

Seal Property of Rubber Cylinder Shoulder in Packer

Li-xin Lu, Tao Deng, Gui-qin Li and Mitrouchev Peter

Abstract The end of the rubber cylinder will flow into the annular space between the central oil pipe and the casing and form, which cause shoulder protrusion. This appearance will cause stress concentration at the end of the rubber cylinder, which may cause the end of the rubber cylinder to be damaged. In order to solve the problem of the shoulder protruding of sealing cylinder of packer, a kind of shoulder protector is designed in this paper. The effect of the designed shoulder protector is verified by the high temperature and high pressure test.

Keywords Packer sealant cylinder · Shoulder protector · High temperature and high pressure test

1 Introduction

Improving the tightness of packer and enhancing the reliability of packer downhole seal is of great engineering significance and practical value for solving the sealing failure of packer in deep wells and ultra deep wells drilling in our country. The research on the tightness of compression packer in mainly concentrates on the mechanical performance test of the sealant cylinder and factors affecting the sealing performance and failure criteria and the sealing structure optimization of packer and high temperature and high pressure test of packer [1]. In optimization of the sealing structure of packers, Zhang Baoling et al. put forward the improved method of packer rubber cylinder under high pressure [2]. The simulation of the packer rubber cylinder anti protrusion device is provided by Zhang Debiao et al. [3]. In the packer

L. Lu · T. Deng · G. Li (✉)
Shanghai Key Laboratory of Intelligent Manufacturing and Robotics, Shanghai University, Shanghai 200072, China
e-mail: leeching@shu.edu.cn

M. Peter (✉)
University Grenoble Alpes, G-SCOP, 38031 Grenoble, France
e-mail: peter.mitrouchev@g-scop.inpg.fr

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*, Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_16

high temperature and high pressure test device, Wang Fengqing design a new type of packer sealing test device [4].

In this paper, the high temperature and high pressure test equipment on the ground is used to simulate the downhole environment of high temperature and high pressure. The sealing performance of the packer is verified, and the sealing structure of the packer is optimized according to the test results.

2 Shoulder Protector Device for Sealing Cylinder of Packer

As shown in Fig. 1, when the packer complete sitting seal, the end of the rubber cylinder flows toward the annular space between the central oil pipe and the casing. The phenomenon of shoulder protrusion causes the stress concentration at the end of the rubber cylinder, which may cause the end of the rubber cylinder to be damaged. In engineering applications, in order to prevent the shoulder protruding, it is usually install the vulcanizing spring at the end or install the red copper wrist guard and other anti protrusion parts. Because the anti protrusion parts and the rubber cylinder are vulcanized together, When creep occurs in the casing, the vulcanized anti-protrusion material is easily separated from the rubber cylinder, and the hardness and modulus of elasticity of the material are greatly different from the rubber cylinder, so that the rubber cylinder may be damaged by the anti-protrusion material during compression.

In order to solve the problem of shoulder protruding of packer sealant cylinder, a kind of shoulder protector is designed. Figure 2 is a new design of a reusable packer shoulder protection mechanism. The mechanism comprises an upper pressing block, a spring, an upper slide block, an upper hoop ring, an upper conical ring, a lower conical ring, a lower Hoop Ring, a lower sliding block and a lower pressing block.

As shown in Fig. 2, the shoulder protector is installed at both ends of the rubber cylinder. The upper slide block and the lower slide block are evenly distributed around the central oil pipe in the circumferential direction. The head of the slide block is arranged in the groove of the pressing block, and the head of the slide block is provided with a spring. The other end of the spring is fixed in the groove of the pressing block. The lower end of the sliding block is a conical surface, and the conical surface is close to the conical surface of the lower conical ring.

When the upper block is subjected to the setting pressure, the upper pressing block moves downward and the head of pressing block is against the shoulders of both sides of the slide block. The sliding block has a tendency to slide along the tapered ring after being pressed by the block. Because the radial displacement of the slider is limited by an aluminum hoop ring, when the radial component of the slider increases continuously, the aluminum hoop is broken and the slider begins sliding

Fig. 1 Shoulder protrusion

along the conical ring toward the casing until the slide block touches the inner wall of the casing. The slide block is evenly distributed around the sleeve, obstructing the annular gap between the central oil pipe and the sleeve, thereby protecting the shoulder of the rubber cylinder and preventing the rubber cylinder from flowing in the annular space. When the packer is unsealed, the pressing block force is removed, the slider reset under the action of the spring.

3 Sealing Test Process and Result Analysis

The experimental procedures are as follows: The temperature is set at 120°, the center cavity pressure is 15 MPa, and the pressure in the upper and lower cavity is 45 MPa. The hydraulic oil is injected into the central tubing and wellbore cavity,

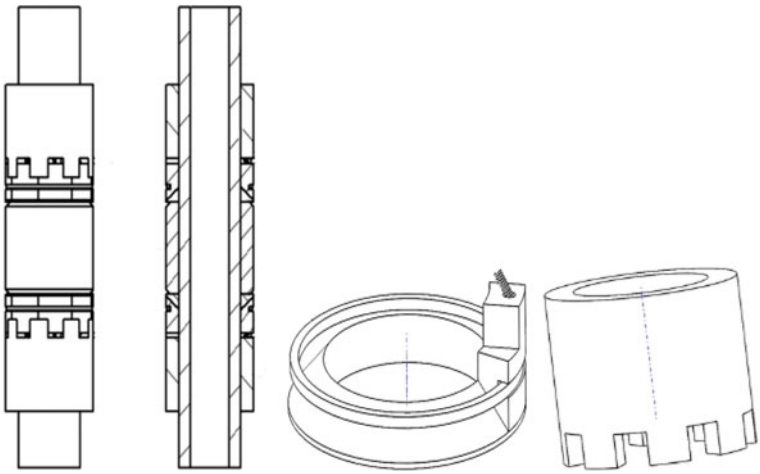


Fig. 2 Shoulder protection

and the pressure is maintained by 5 MPa for 5 min to check whether the wellbore leaks. If the shaft not leaking, the casing is heated to 120° and preserving heat in a temperature controller. Increase the central pipe pressure to 15 MPa. After maintaining the pressure for 5 min, the packer reaches a sealing state and starts to test the tightness of the upper cavity. Releasing lower cavity pressure, increase the upper cavity pressure to 45 Mpa. After maintaining the pressure for 20 min, Re-open the central pipe pressure, keep the pressure for 5 min, turn off the upper cavity pressure, open the lower cavity pressure about 45 MPa, and maintain the pressure more than 20 min.

There are the curve graph for the sealing test of the optimized size of the sealant cylinder and the original sealant cylinder. The sealing test curve of the original rubber cylinder is shown in Fig. 3. The packer does not have a rubber cylinder shoulder protector, and the sizing of the rubber cylinder has not been optimized. The packer tested in Fig. 4 uses a rubber cylinder shoulder protector and the sizing of the rubber cylinder has been optimized.

As can be seen from Fig. 3, when the central pipe is pressurized to 15 MPa and lasts 5 min, the packer reaches sealing state. The fluctuation of the upper cavity pressure in the curve is basically stable at about 45 MPa, but in the lower cavity pressure test, the fluctuation of the pressure curve is obvious. As can be seen from Fig. 4, the pressure curve of the upper cavity fluctuates somewhat, but basically remains stable. In 65 min, the pressure in the curve cavity is approximately stable at about 45 MPa, and the fluctuation amplitude is lower than the original packer's.

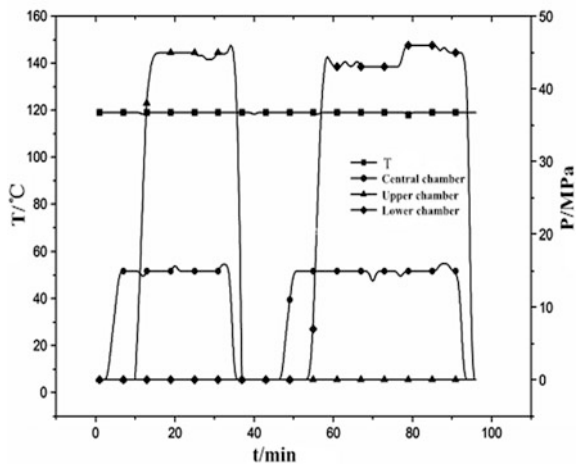


Fig. 3 Sealing test curve of original

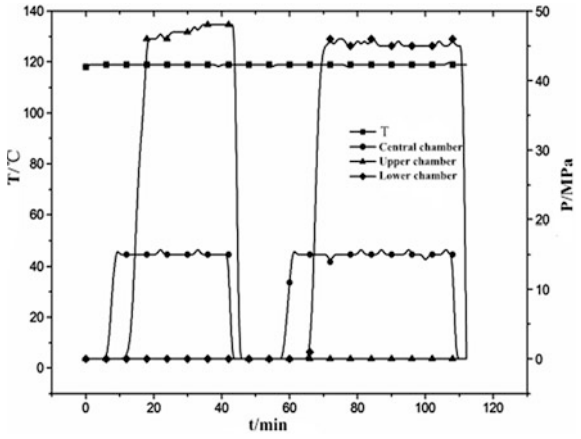


Fig. 4 Sealing test curve after optimization rubber cylinder

In order to test the protective effect of shoulder protection device of packer rubber, Extensive tests have been carried out with a shoulder protection device of the packer and without shoulder protection device of the packer separately. As shown in Fig. 5. Without the protection of the packer, there is a breakage near the shoulder of the rubber cylinder, which indicates that the shoulder of the sealing component is torn during the sealing process. As shown in Fig. 6, it can be seen that the packer with a shoulder protector is basically intact and has no obvious damage on the surface.

Fig. 5 Test results with shoulder protection

Fig. 6 Test results without shoulder protection

4 Conclusion

The shoulder protector can effectively reduce the shoulder protrusion of the rubber cylinder and makes the stress distribution of the contact area more uniform. The experimental results show that the shoulder protector of the sealing component can effectively protect the sealing rubber cylinder of the packer, improve the service life of the sealing component, and simultaneously determine the shoulder protrusion of the rubber cylinder is an important factor leading to the tearing of the rubber cylinder.

Acknowledgements This research is partly supported by the research program of Shanghai Science and technology committee (NO.14DZ1204203).

References

1. Jian Chen, Boyuan Tian, Yuwen Liu (2010) Failure analysis and measures of compression packer rubber cylinder. J Sci Technol Inf 31:80–84
2. Chang Z, Xilu W, Xu XP (2009) Improvement of sealant cylinder for high pressure packer. J Pet Field Mach 01:85–87
3. Debiao Zhang, Jun Hou, Wei Ma et al (2010) Influence analysis of anti outburst structure on sealing performance of packer. J Pet Mach 12:53–55
4. Shaodong Ju, Renqi Ma, Xiumei Hong et al (2015) Design and sealing performance analysis of offshore completion packer rubber cylinder. J Pet Mach 06:46–49

Data Acquisition and Storage Network Based on CAN-Bus

Lixin Lu, Jun Hua, Guiqin Li and Mitrouchev Peter

Abstract A strong versatility, scalability of distributed multi-channel data acquisition and storage network is put forward in this paper. A multi-channel, real-time data acquisition and storage module is designed based on the research of pressure sensor and A/D conversion. Then the bidirectional and multi-station communication system is constructed bases on building the data acquisition and storage module on the CAN-bus. The whole system can be synchronized for multi-point data acquisition, and store data for data traceability.

Keywords Data acquisition · Data storage · CAN-bus

1 Introduction

The detection system is required to adapt to multi-target, distributed, scalable in the highly automated, complex tasks of the automated production line. Haifei Ding et al. develop a multi-channel data acquisition and storage system based on ADS8365, which couldn't apply to distributed detection. It was a single module and couldn't constitute a multi-node communication network [1]. Aboli Audumbar Khedkar develops a high speed data acquisition system based on FPGA, which focuses on high-speed data acquisition and lossless storage but not for the industrial environment. It couldn't transmit data through the field communication network [2].

This paper sets the intelligent detection as the target, realizing decentralized data acquisition for a number of detection points. Developing the system consists of the

L. Lu · J. Hua · G. Li (✉)
Shanghai Key Laboratory of Intelligent Manufacturing and Robotics,
Shanghai University, Shanghai 200072, China
e-mail: leeching@shu.edu.cn

M. Peter (✉)
University Grenoble Alpes, G-SCOP, 38031 Grenoble, France
e-mail: peter.mitrouchev@g-scop.inpg.frn

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_17

ARM processor as the core [3], and the CAN-bus integrated as the communication network for distributed data acquisition and storage.

2 Overall Scheme Design

The data acquisition and storage network system based on the CAN-bus is mainly composed of two parts. The first part is the terminal sensing unit and STM32F103 chip integrated as the data acquisition and storage module. The second part is the communication network based on the CAN-bus with the core of STM32F407 chip. Each data acquisition and storage module is a node of the communication network, as shown in Fig. 1.

The workflow of system detection and data acquisition and storage is shown in Fig. 2.

3 Data Acquisition

The system aimed at the pressure signal detection of automated production line, uses balloons and pressure sensor as the terminal sensing module to detect the relative pressure within the balloon as the data source. The system is designed to convert any gas pressure signal into 0–3.3 V voltage signal, and then the digital value range 0–4095 is obtained by 12-bit A/D conversion. In order to ensure that the system has a considerable sensitivity to the signal source, it is fundamental that the sensor ranges 0–50 kPa, and the sensitivity is 100 mV/kPa, is shown in Fig. 3.

The divider circuit is calculated as follows: $V_{\mathrm{AIN}} = V_S \times \frac{R_0}{R_0 + R_1}$, and $V_{\mathrm{AIN}} \leq V_{\mathrm{REF}+} \leq 3.6V$, $I_{O+} = V_S/(R_0 + R_1)$.

In the formula: V_{AIN} is the input voltage of the analog-to-digital converter; $V_{\mathrm{REF}+}$ is the system circuit reference voltage; V_S is the output voltage of the pressure sensor; R_0, R_1 is the resistance for voltage divider; I_{O+} is the sensor's full-scale output current.

The standard output voltage of each sensing unit is consistent by using the same voltage divider circuit. And then the stability and reliability of the data collected by the system is ensured.

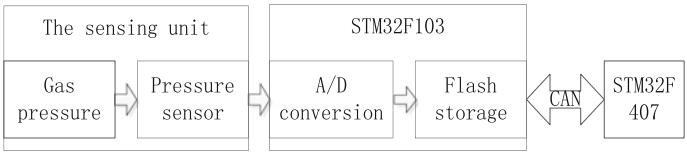


Fig. 1 Data acquisition and storage block diagram

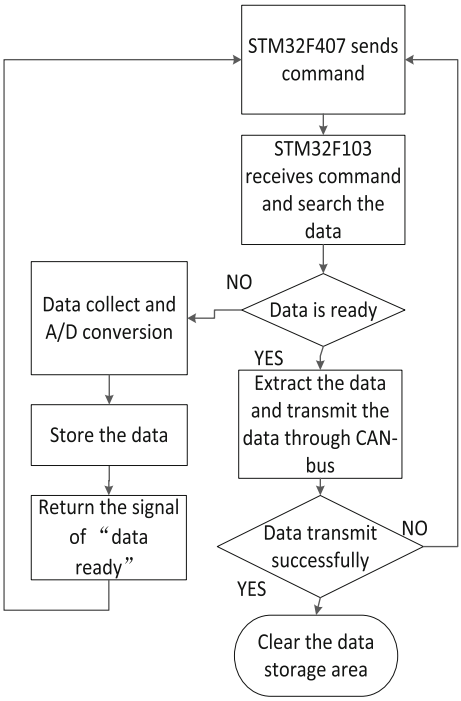


Fig. 2 The workflow of system detection and data acquisition and storage

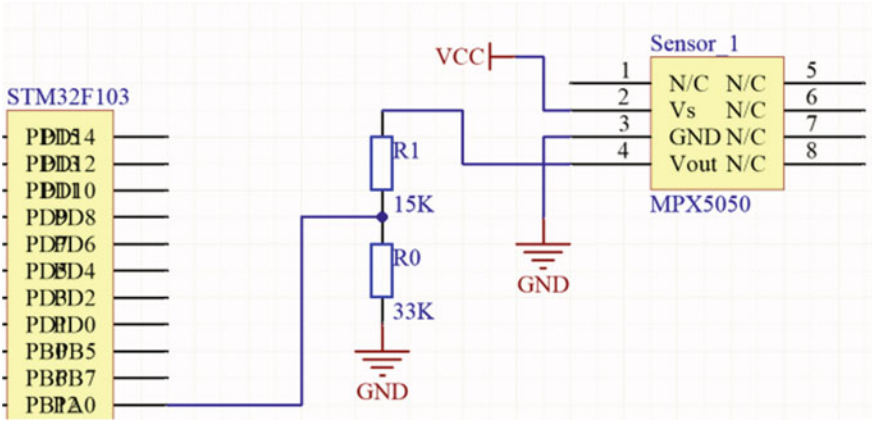


Fig. 3 The voltage divider circuit

The system sets the single-channel signal sampling frequency of 100 Hz, and the overall sampling frequency is 1200 Hz as each A/D converter is responsible for twelve A/D conversion channels.

The successive comparison is adopted as the A/D conversion mode in order to obtaining higher conversion rate under lower power consumption.

The system uses 12-bit high-precision A/D converter with a total of twelve analog input channels: PA0-PA7, PC0-PC3. According to the A/D conversion resolution formula: $\Delta U/2^n$, the A/D conversion resolution of the system is 8 mV. In the formula, ΔU means the amount of the change of the input analog voltage, n means the A/D conversion's bit number.

4 Data Storage and Communication

This paper proposes a sequential storage mode. Each detection collects amounts of data of sensors needed, and then determines the buffer of the data according to the pointer. At the same time, the pointer matches the attribute information of the cached data in the form of a list, and each detection data is sequentially cached and the pointer value is recorded.

The communication instruction sent to STM32F103 contains the function code, parameter code and the attributes of data collected including the ID of the sensor, the amount of data needed. The system traverses the query of all the data storage areas to realize data storage and extraction.

This paper introduces CAN-bus as the framework of the system's communication network which is easy to update and expand the terminal sensing module [4]. The system prioritizes the data packet so that the data packet can be transmitted on the communication bus in sequence as well as avoiding communication bus congestion. The bulk of the sensor data is divided into eight bytes as a group of data packets to be transmitted. And each data packet will be checked by CRC (Cyclic Redundancy Check) so that the stability and reliability of data transmission can be improved effectively [5]. Data acquisition and storage process is shown in Fig. 4.

5 Data Analysis and Verification

This section aims at validating that the system has achieved the target through extracting and analyzing the data collected. The waveform of the data collected directly by the system is shown in Fig. 5. The abscissa in the figure is the number of data points collected and the ordinate is the gas pressure value obtained by A/D digital conversion. It can be seen from the figure, the effective signal characteristics are the three peaks whose differences of the pressure reach 10 kPa, which means the system provides a reliable pressure signal peak and the reference value. The abnormal peak might appear when the pressure value in the original data is small with the loud signal noise. The data smoothing process is adopted in STM32F407 in order to improving the signal-to-noise ratio of the signal collected and highlight the signal amplitude characteristic.

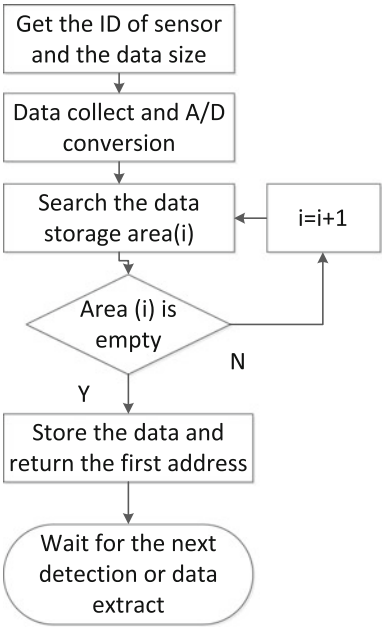


Fig. 4 Data acquisition and storage process

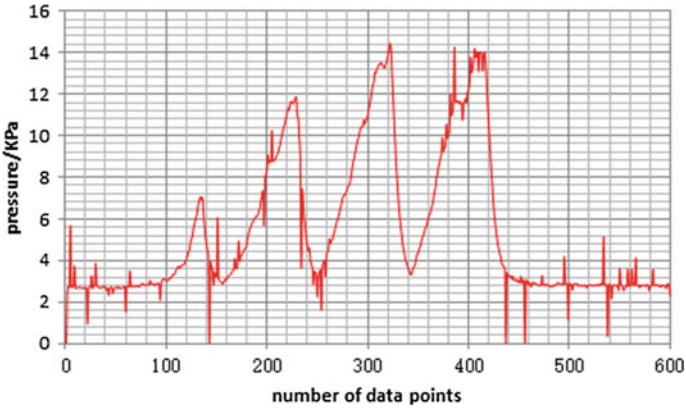


Fig. 5 Raw data waveform

The system uses the sliding arithmetic mean filtering method with 8 data points for a width to filter out the clutter signal. The processing data waveform shows in Fig. 6. The noise signal is significantly reduced after the smoothing process. The effective peak and reference value can be clearly obtained, and the miscarriage of judgement caused by the noise signal is avoided.

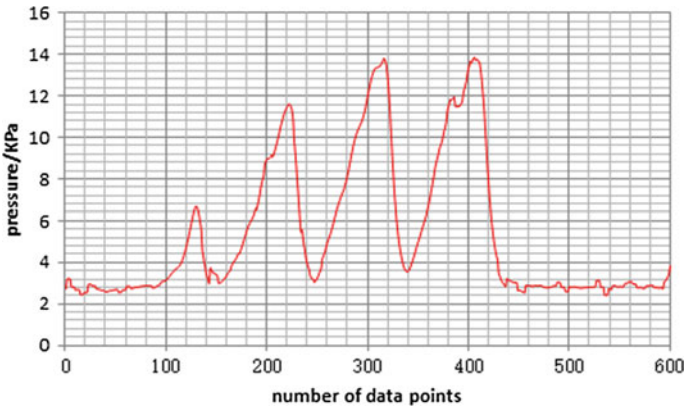


Fig. 6 Data waveform after smoothing process

6 Conclusion

This paper put forwards a new data collection mode of the distributed system, in which the data acquisition and storage directly correspond to the terminal sensor. The function of collecting and storing different data source is realized, which brings the commonality of the system in the field of distributed data collection and storage. The new communication mode is introduced into the data acquisition and storage system, setting eight bytes as a data packet, which is short in length and fast in transmission. It is not only convenient for the internal instruction, but also for ensuring the stability of transmission of large quantities of data [6].

Acknowledgements This research is partly supported by Shanghai Automotive Industry Science and Technology Development Fund (NO.1512).

References

1. Khedkar AA, Khade RH (2017) High speed FPGA-based data acquisition system. Microprocess Microsyst 49:87–94
2. Ding HF (2011) The design of multiplex data acquisition and storage system based on ADS8365. IEEE Proceedings of 2011 4th IEEE international conference on computer science and information technology, vol 4. pp 643–646
3. Chenhui W, Yue W, Kai Y (2016) Design of multi-channel data acquisition system based on STM32. Meas Control Technol Instrum 1:51–53 + 57
4. Wang H (2006) Information system about the CAN bus. China Meas Technol 32(1):130–131 + 144
5. Jiqiang Xia, Chuansen Zhang, Ronggang Bai, Liqiang Xue (2013) Real-time and reliability analysis of time-triggered CAN-bus. Chin J Aeronaut 01:171–178
6. DaoGang P, Hao Z, Hui L, Kai Z (2009) Design and implementation of ARM embedded monitoring platform based on modbus protocol. Electr Power Autom Equip 01:115–119 + 123

Effect of Pressure on Packer's Sealing Performance

Li-xin Lu, Lei Tang, Gui-qin Li and Peter Mitrouchev

Abstract The sealing performance of packer under different setting pressure is studied in this paper, and the finite element analysis is carried out. Firstly, the finite model of packer is established by using ANSYS software. Different setting pressure are provided, which cause the compression change of packer rubber. The relationship between the setting pressure and the compression distance and the maximum contact stress of the packer rubber is found out by analyzing this change, including the gap between the rubber and the central pipe and the contact stress between the rubber and the inner wall of the casing. Finally, the observed phenomena and simulation data are analyzed and summarized, and the influence of the setting pressure on the sealing performance of the packer is obtained.

Keywords Setting pressure · Packer · Sealing performance · Finite element analysis · Rubber

1 Introduction

Well completion engineering is an important part of oil drilling and production, directly relates to the service life and output of oil well [1]. Compression packer is a kind of important oil well completion tool. It mainly plays the role of separating different production layers, fracturing, acidification, water injection and so on. Sealing performance is the core property of packer [2], Rubber is the core part of packer, and it is the working part of packer, which is directly related to well completion quality and underground safety. RIVENBARK M, et al. designed a

L. Lu · L. Tang · G. Li (✉)
Shanghai Key Laboratory of Intelligent Manufacturing and Robotics, Shanghai University, Shanghai 200072, China
e-mail: leeching@shu.edu.cn

P. Mitrouchev (✉)
University Grenoble Alpes, G-SCOP, 38031 Grenoble, France
e-mail: peter.mitrouchev@g-scop.inpg.fr

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_18

DHM-type compression packer. The packer with double sealing components, hydraulic sealing, mechanical releasing [3]. With the sealing performance of ultra-deep well RTTS packer analyzed, Tong Shaokai put forward packer sealing failure criterion [4].

The study on the creep and stress relaxation of packer rubber provides the basis for the optimization of packer sealing structure. In this paper, the optimum seating load is obtained, owing to finite element analysis of the compression distance and maximum contact stress of packer.

2 Finite Element Model

The geometrical model of the packer sealing structure is established in ANSYS [5]. In the mesh generation, the sealing rubber is the main object of the calculation and analysis. As shown in Fig. 1, polygon mapped mesh generation is used. Non-critical components, such as Center tubing, casing, base, use free meshing generation. The number of the grid nodes is 1766, the number of units is 511, and the average value of the unit quality is 0.9424. As the cement bond on the outside of the casing, the casing is fixed, the central oil pipe, the pressing block and the base are provided with radial displacement constraint, and the base is bound with the central oil pipe.

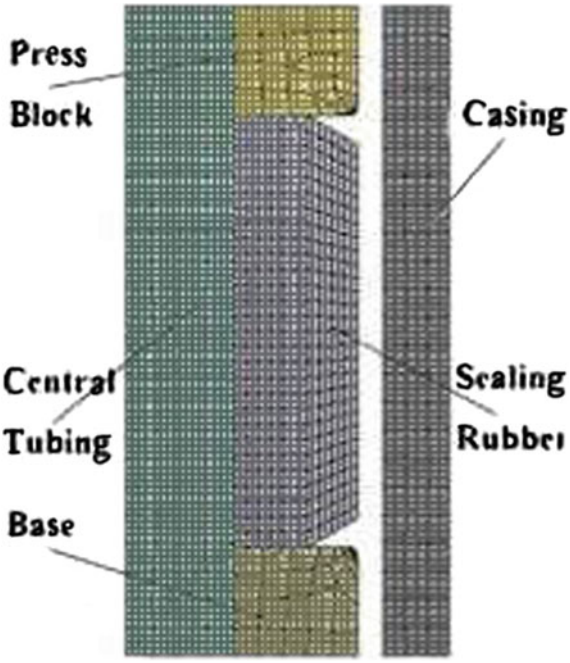


Fig. 1 Meshing diagram of sealing structure

3 Analysis on Sealing Performance

When the compression packer is positioned below the preset position, the pressure is injected through the central pipe to compress the rubber so as to realize the action of sealing. Reasonable setting pressure can ensure the sealing performance of packer [6]. Too little setting pressure does not complete the sealing or incomplete sealing, and too large setting pressure will cause damage to the rubber. Therefore, it is necessary to analyze the effect of different setting pressures on the sealing performance of the packer.

When the setting pressure reaches 12 MPa, the amount of deformation of the packer sealing rubber is shown in Fig. 2, and the contact stress between the sealing rubber and the casing is shown in Fig. 3. Caused by the increasing of setting pressure, the deformation reaches 27.748 mm, simultaneous the maximum contact pressure reaches 3.4978 MPa. At this point, the gap between the rubber and the center tube is completely disappeared and the rubber has been basically compacted, which means that the packer completes the sealing. And the annular gap between the center oil pipe and the casing is sealed.

The distribution of the contact pressure between the sealing rubber and the inner wall of the casing under different loads is shown in Fig. 4. In pacing with the setting pressure increases, the stress in the contact region rises as a whole. When the setting

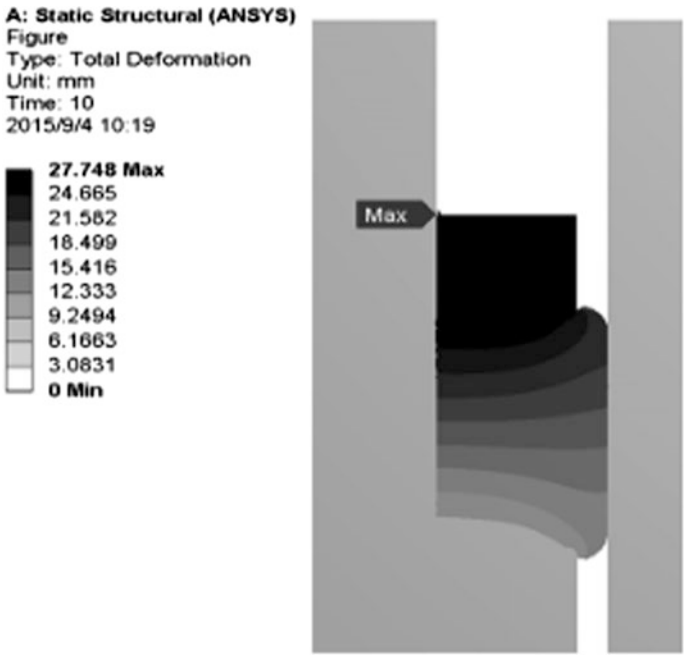


Fig. 2 Total deformation

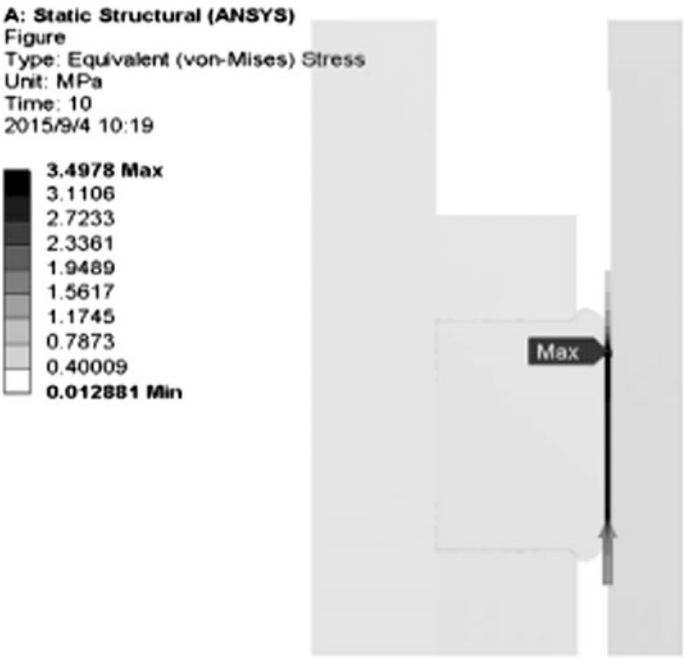


Fig. 3 The contact stress distribution

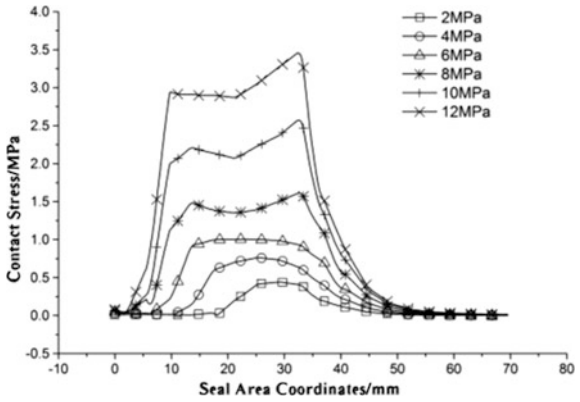


Fig. 4 Contact pressure distribution

pressure is between 2 and 6 MPa, the contact stress curves indicate that the rubber is always in the gap with the central tubing and the rubber is not fully contacted with the inner wall of the casing, which are single peak curves. When the setting pressure is small, the radius of curvature of the rubber bends towards the central tubing, but the rubber begins to bend in the direction of the casing after contact with the central tubing with the increasing of setting pressure. Due to the limitation of the

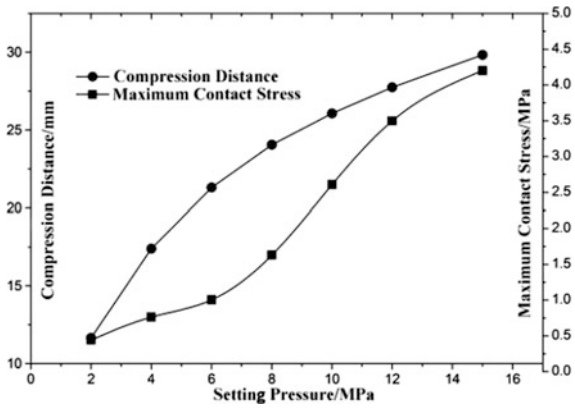


Fig. 5 Effect of setting pressure

central oil pipe, the rubber produces two bends between the upper and the lower, and the two bends simultaneously point to the inner wall of the casing. Precisely because of this, when the setting pressure is more than 8 MPa, the curves have two stress peaks.

Figure 5 shows the change of the compression distance and the maximum contact stress of the packer sealing rubber under different setting pressure. As can be seen from the diagram, with the increase of the setting pressure, the compression distance of the sealing rubber is increasing, but the slope of the curves is decreasing. After the stress reaches 10 MPa, the growth curves of the compression distance begin to become gentle. This is because, as the pressure increases, the rubber is constantly compacted and the compressibility decreases slightly. In the contact pressure curves, the curves grow slowly in the range of 0–6 MPa. And within 6–10 MPa range, the slope increases and the rate increases rapidly. When setting pressure exceeds 14 MPa, the contact pressure increases gradually. After the setting pressure reaches 15 MPa, it is also shown in the figure, the compression distance is almost stable at around 30 mm, and the contact stress is between 4 and 5 MPa. At this point the packer has reached the sealing requirements, can achieve a stable seal.

4 Conclusion

In this paper, the finite element model of the sealing parts is established according to the actual size of the compression packer and the experimental material constants. The simulation results show that the optimal setting pressure of the packer is 15 MPa, and the limit value both of the compression distance and the contact stress are 30 mm and 4.25 MPa respectively.

Acknowledgements This research is partly supported by the research programmer of Shanghai Science and technology committee (NO.14DZ1204203).

References

1. Ju SD, Ma RQ, Hong XM et al (2015) Analysis of sealing performance and design of packer rubber for sea well completion. Petrol Mach 06:46–49
2. Bu YH, Ma MX, Li JH et al (2011) Study on the sealing criterion and structural design method of packer. Lubr Seal 11:75–78
3. Rivenbark M, Dickenson RW (2011) New open hole technology unlocks unconventional oil and gas reserves worldwide: proceedings of the SPE Asia Pacific oil and gas conference and exhibition. Society of Petroleum Engineers
4. Kai TS (2014) Mechanical analysis and control technology research on super-deep high-pressure gas wells packer. Xi'an university of petroleum
5. Dou YH, Ma ZH, Chen J, et al (2015) The finite element analysis on the mechanical behavior of THT packer rubber. Oil and Gas Well Test 03:4–7–75
6. Slay B, Webber W (2011) Stress relaxation of elastomer compounds. Sealing Technol 11(2):9–12

Acquisition and Control Systems of Distributed Data Based on STM32

Lu Lixin, Shen Deshuai, Li Guiqin and Mitrouchev Peter

Abstract A wireless distributed data acquisition and control system which is suitable for Industrial monitoring and control areas is present in this paper. The upper computer of the system is programmed by using LabView, and it communicates with lower computer via WIFI. A STM32F407 card as main control card, and three STM32F103 cards as slave card, contained in lower computer system, in which the cards communicate with each other via RS485. Slave cards are used to collect data and control peripherals. This system has a good many advantages: high transmission rate, strong anti-interference capability, simple wiring, easy to extend etc., and it apply to functional detection device of massage chair successfully.

Keywords Embedded system · Data acquisition · WIFI communication

1 Introduction

Wireless transmission system based on embedded is a hot research spot in industrial automation, measuring, controlling, electrical appliance, medical treatment areas, and has wide application [1–3]. Gang Peng et al. propose a data acquisition system based on STM32 and RS485 communication, yet the communication between PC and lower computer is through USB [4]. Hong Gang Li et al. design the system to receive data and analyse wave via LabView [5].

This paper proposes a data acquisition and control systems based on STM32 and LabView, apply to detect the performance of massage chair online, and realize long-range distributed data acquisition and peripheral control.

L. Lixin · S. Deshuai · L. Guiqin (✉)
Shanghai Key Laboratory of Intelligent Manufacturing and Robotics,
Shanghai University, Shanghai 200072, China
e-mail: leeching@shu.edu.cn

M. Peter (✉)
University Grenoble Alpes, G-SCOP, 38031 Grenoble, France
e-mail: peter.mitrouchev@g-scop.inpg.fr

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_19

2 Hardware Design

STM32 microcontroller is ST's 32-bit microcontrollers based on ARM Cortex-M3 core designed for high performance, low cost, low power embedded applications. The STM32F103 card used in our system has three 12-bit ADC and more than 11 timers, extended 485 communication module. STM32F407 is main control board and it has a higher operating frequency, including WIFI communication module. Using differential signal and twisted pair to achieve longer distance transmission, RS485 communication has stronger anti-interference ability. Because of these advantages, it is more suitable for distributed systems, thus all the cards in our system communicate with each other by RS485 communication.

The system block diagram is shown in Fig. 1. The lower computer contains a STM32F407 master board and three STM32F103 slave boards, and each of the three slave boards has its own extension board. The two slave cards are used for collecting data, and the external data are collected by a pressure sensor, then analog signal is connected to pin of the slave board through the voltage dividing circuit of the expansion board. After the data being AD converted, it is transmitted to F407 card through 485bus. By amplifying circuit of expanding board, third slave card is used to control the stepper motor and solenoid valve and other peripherals. The peripheral control circuit is shown in Fig. 2.

The control system includes a variety of peripheral control, including relay controlled tricolor light, stepper motor, etc. To achieve these controls, isolation and

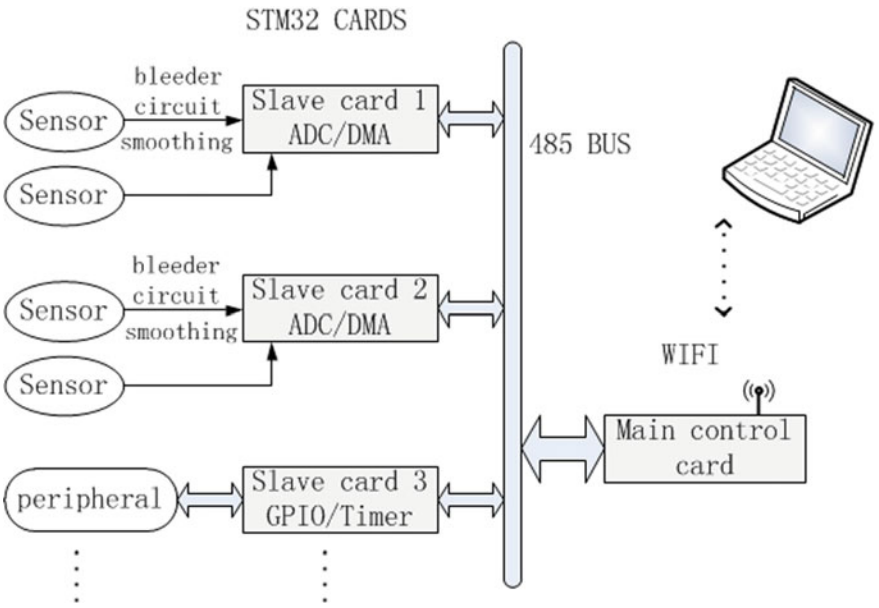


Fig. 1 System block diagram

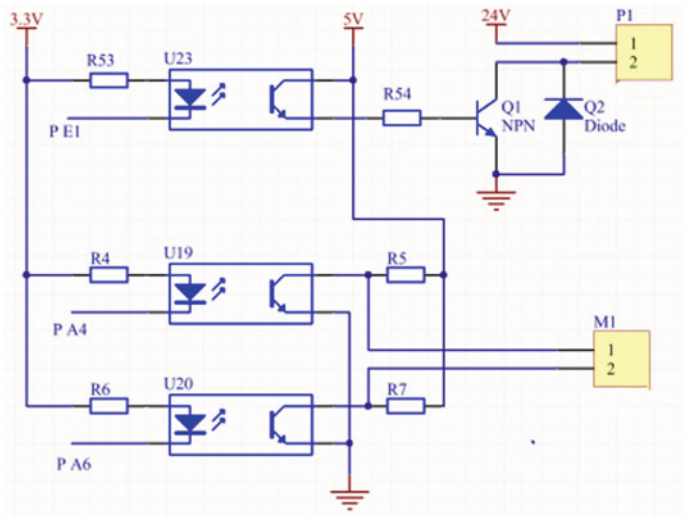


Fig. 2 Peripheral control circuit

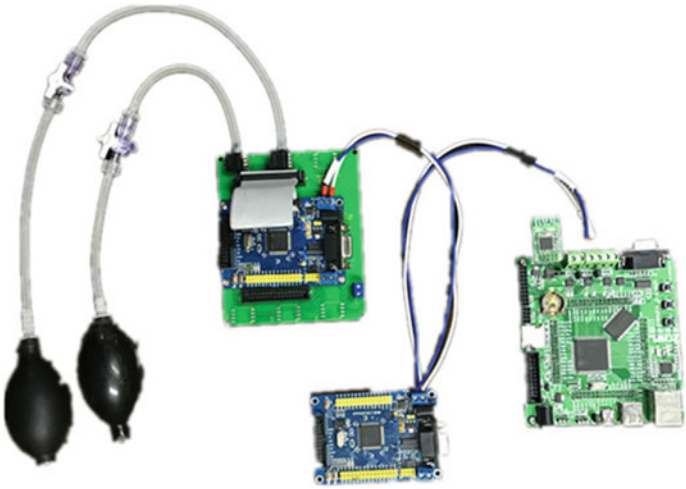


Fig. 3 Part of electrical material

amplification circuits are designed for third blocks from the board. The relay is controlled by pin PE1, and the isolated circuit of the optocoupler controls the on-off of the triode, and then realizes the on-off of the relay. The stepper motor interface contains pulse and directional signals, and the driver is common anode 5 V control.

As Fig. 3 shows, two black airbags connect sensors through the three-way valve. The STM32F103 is connected to the expansion board by line. Different expansion boards can control different peripheral devices and realize different functions.

3 Software Design

Figure 4 is the software flow chart of the control system, including the flow chart of the host computer and the STM32F407 master board. After receiving the control command, two slave 103 cards collect data from the corresponding channel from the board and its sampling frequency is 100 Hz. The third slave card controls the relay through the output pin, or enables the timer to carry out the PWM output, and controls the stepping motor.

The RS485 bus simply defines the physical layer of the network and does not specify the upper level protocol. Therefore, our system uses Modbus protocol which is widely used in industry to communicate with master slave device. TCP/IP protocol is used between the master board and the host computer to communicate with WIFI. In order to facilitate multiple PC access, the master board is used as the server to create AP and wait for the host computer to access.

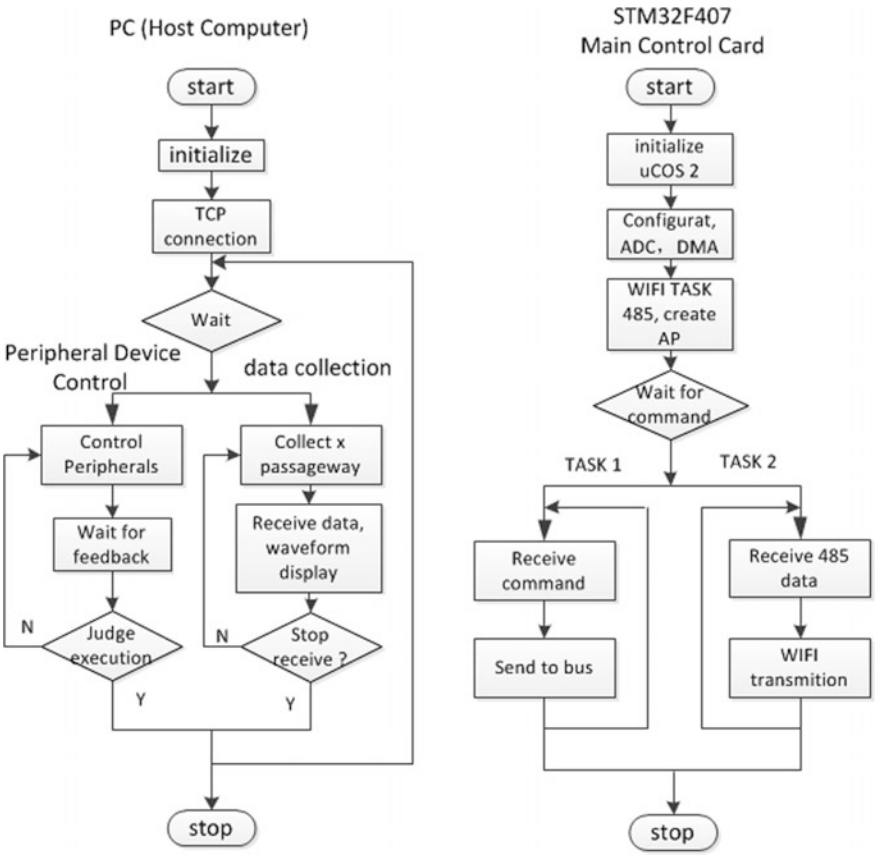


Fig. 4 Software flow chart

The host computer is written in LabVIEW, which is a graphical programming tool for NI Company. Firstly, initialize the control and create the connection with the master board based on the IP address and port number. The host computer receives data from main card and extracts the amplitude and frequency information through the single frequency measurement and other nodes after the waveform is extracted.

4 Results

The system includes PC host computer and four embedded boards. The two plates collect the pressure in the four airbags respectively, and the third board is used to control tri-colour lights and stepping motors. Figure 5 is the upper computer operation interface, used to control the lower computer. Figure 6 is the Waveform

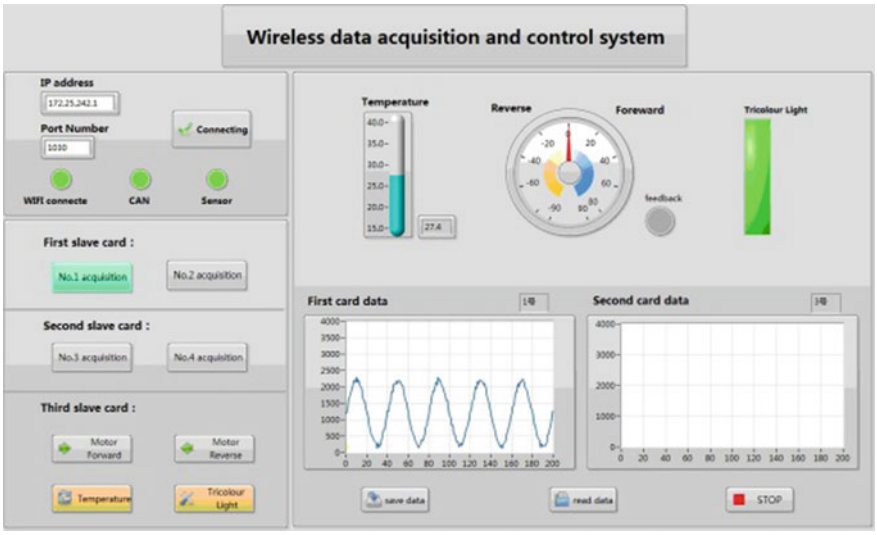
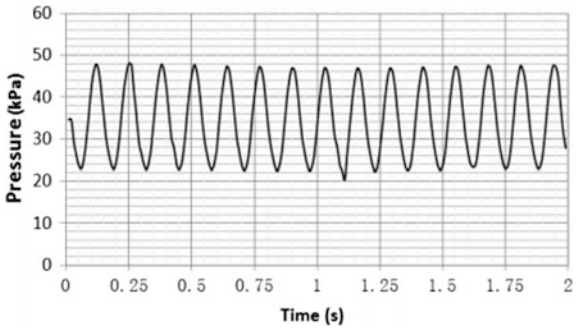


Fig. 5 Operation interface

Fig. 6 Waveform acquisition



collected by the system when the airbag is squeezed at the frequency of 7.5 Hz. Ordinate indicates the value of pressure, changing from 20 to 50 kPa.

5 Conclusions

The wireless distributed system proposed in this paper can realize the functions of data acquisition and peripheral control. Meanwhile, due to the use of 485 communication, the lower computer can be extended conveniently, and then realize the multi node control. This system can be widely used in industrial measurement and control and other fields.

Acknowledgements This research is partly supported by Shanghai Automotive Industry Science and Technology Development Fund (NO.1512).

References

1. Gang P, Xu QJ, Zhang CG, Wang ZN (2011) Design of RS485 bus distributed data acquisition system based on STM32 MCU. J Servo Control 02:64–67 + 84
2. Zhang X, Yuan XG (2010) Design of power data acquisition system based on STM32. J Electron Meas Technol 11:90–93
3. Li Honggang, Zhang Suping (2014) Design of multichannel data acquisition system based on single chip microcomputer and LabVIEW. J Foreign Electron Meas Technol 04:62–67
4. Zeng L, Zhang HF, Hou WY (2011) Design and implementation of wireless measurement and control system based on WIFI. J Electr Meas Instrum 07:81–83 + 96
5. Zhang TS, Li YM (2012) Data acquisition design based on the STM32. J Adv Mater Res 591–593:1527–1530

Introduction of Cyber-Physical System in Robotized Press-Brake Line for Metal Industry

Beibei Shu, Gabor Sziebig and Bjørn Solvang

Abstract Bin picking is a typical work, which is easy to automate up to a given complexity of the work-piece dimensions. In case of casted work-pieces, the dimensions are most of the time not accurate enough for an industrial robot to be able to pick it up without additional sensors/intelligence. In this paper we introduce a cyber-physical system, where all sensors, actuators, machines and industrial robot is connected to a local network, where they share information easily with each other. The novelty of the system that the proposed solution is achieved on the software side, with minimum hardware reconfiguration need. We keep the flexibility of the industrial robot, but extend its' understanding with sensor fusion on a higher decision level, rather than on low robot programming level. The system also allows remote monitoring and supervision of the production plant.

Keywords Industrial robot · Industry 4.0 · Cyber-physical system
ROS

1 Introduction

In recent years, industry 4.0 has become a very popular topic in industry-related fairs, conferences, or call for public-funded projects. The concept of industry 4.0 was to develop from Internet of Things, which was created by British technology pioneer Kevin Ashton in 1999 [1]. Then in April 2013, the Industry 4.0 Working Group present a final report to the German federal government. The report categorizes human industry into four stages, from steam engine stage—first industrial revolution, electrically-powered mass production stage—second industrial revolution and electronics-information stage—third industrial revolution, to cyber-physical based production stage—fourth industrial revolution [2]. The fourth stage is the Industry 4.0 which includes cyber-physical system, Internet of things

B. Shu (✉) · G. Sziebig · B. Solvang
UiT, Campus Narvik, Lodve Langesgate 2, 8514 Narvik, Norway
e-mail: beibei.shu@uit.no

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_20

and cloud computing [3]. As for now, we are in the industrial revolution but not finish this revolution. Germany is planning to promote the revolution process and Germany is not the only country planning to upgrade their industry. The industry upgrading becomes a trend across the world recently, the representative countries' plan including: Japan—Super Smart Society [4], Norway—Norge 6.0 [5], United States—The Advanced Manufacturing Partnership [6], China—Made in China 2025 [7].

Cyber-physical systems (CPS) are enabling technologies, which bring the virtual and physical worlds together to create a truly networked world in which intelligent objects communicate and interact with each other [8]. In this paper, we will be introducing a solution for a robotized press-brake line, where all system are connected to the network (a true cyber-physical system) and the progress of the production can be either supervised on-site, in a web browser or in a simulation. This is achieved with the following tools: system based components are connected to the network with the usage of Robot Operating System (introduced in Chapter "Application of Long Short-Term Memory Neural Network to Sales Forecasting in Retail—A Case Study"), supervision of the robotized solution is solved by a software solution (called FlexGui 4.0), which is also introduced in Chapter "Application of Long Short-Term Memory Neural Network to Sales Forecasting in Retail—A Case Study". In Chapter "", the application of the solutions mentioned above will be presented, while Chapter "A Coarse-to-Fine Matching Method in the Line Laser Scanning System" provides the conclusion.

2 Software Based Control

The Robot Operating System (ROS) is an open source, multi-platform robot controlling software, which is a collection of variety tools and libraries [9]. ROS is widely used among the researchers, but it is also developed together with industry and is highly promoted between industrial partners, as all results that are achieved in research can be implicitly implemented in industrial version. In ROS, there is a very basic concept—Nodes, which are processes that perform computation. For example, one node controls a servo motor, one node control optical sensors and one node control electro-pneumatic valves, see in Fig. 1. Multiple nodes can run in a same ROS device, and multiple ROS devices can connect together. In between all the connected ROS device or inside a ROS device, each node can communicate with each other by passing Messages, which is simply a data structure. Under nodes, there are Topics and Services and a node can contain many topics and services. A topic is like a message bus, see in Fig. 2. You can get messages from a subscribed topic if someone publishes data in it, or send messages to other parties by yourself. A service is like a function, see in Fig. 3. You can call it to run a specific script and wait for a call-back function, but the service can only response one request each time.

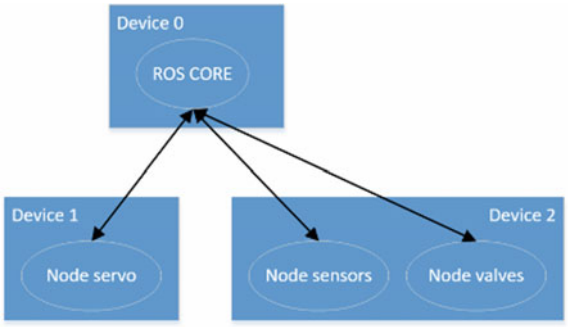


Fig. 1 ROS node

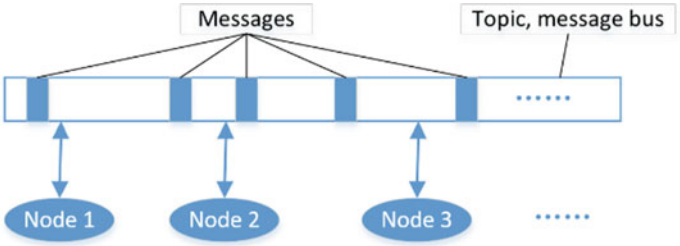
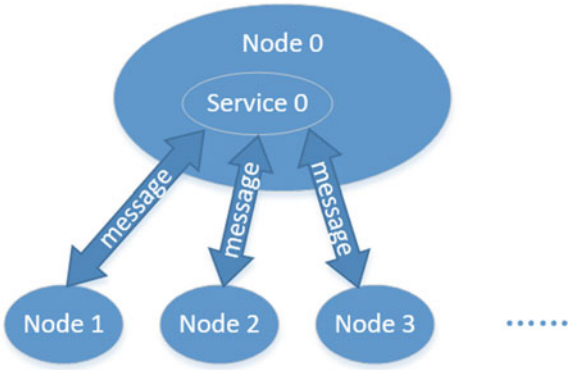


Fig. 2 ROS topic

Fig. 3 ROS service



FlexGui 4.0 is an advanced open source user friendly industrial robot controlling interface, which is based upon popular web technologies enabling the possible to run it on almost every modern browser supported device. The user will have exactly the same user experience on each device. FlexGui 4.0 is used to show information and send user inputs to the server, which is the ROS device. Since FlexGui 4.0 is

based on ROS, after connected to ROS, FlexGui 4.0 will display all the nodes in the system, and offer a graphic interface for user to read or set variables. For customized variables, user can publish them to topics or send request to services. Based on this mechanism, FlexGui 4.0 is not suit for real-time control functions. While, it is possible for ROS connected to multiple FlexGui 4.0 clients, each FlexGui 4.0 client will synchronize with the server and show the same project, even if it changes only one of them. This makes it is possible for users to develop a project simultaneously [10].

3 Application

Gazebo is a well-designed free robot simulator, and has already embedded into ROS. When the Gazebo is running, ROS server will generate a ROS node called "gazebo". Meanwhile, FlexGui 4.0 is connect to ROS server, and it will recognize the node gazebo. User can change a robot-moving project without interfere the process line, and test it virtually in Gazebo. If the project passes the test, user can connect physical robot with ROS server, conduct actual test or introduce it into processing line. When the physical robot is running, user can see a simulated robot additionally in Gazebo, Fig. 4.

In Fig. 5, we can see a virtual press-brake line in Gazebo and a robot control interface in FlexGui 4.0. During the simulation or real manufacturing process, the Gazebo will show the robot's activity and FlexGui 4.0 will offer a detail state of each sensor. And for each element on the FlexGui 4.0 interface such as button in this case, we can change the number, the name, the location, the size or the background function of the button. Multiple robot instructions can be integrated in one button. So, the robot will execute a series of actions just after one button clicked.

In Fig. 6, that is a real case in physical world using this ROS, FlexGui 4.0 and Gazebo solution. Work-pieces that are coming from the moulding machine, robot picks these up and sets on the press-brake, then two pieces running the production

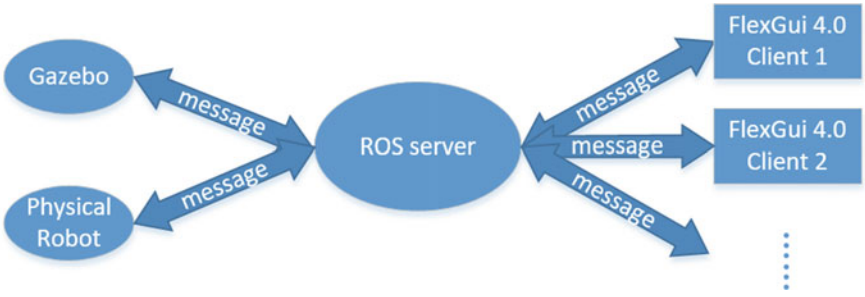


Fig. 4 Connect gazebo and physical robot

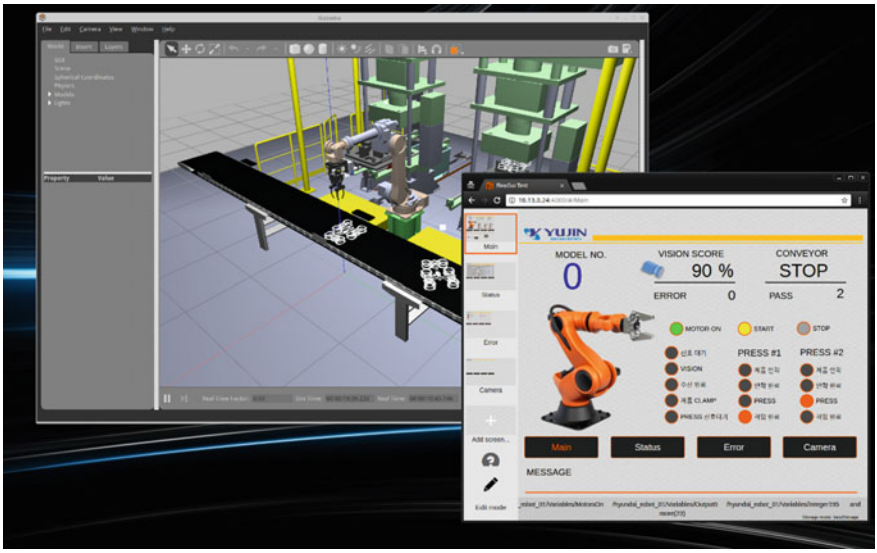


Fig. 5 Gazebo simulation and FlexGui 4.0 interface

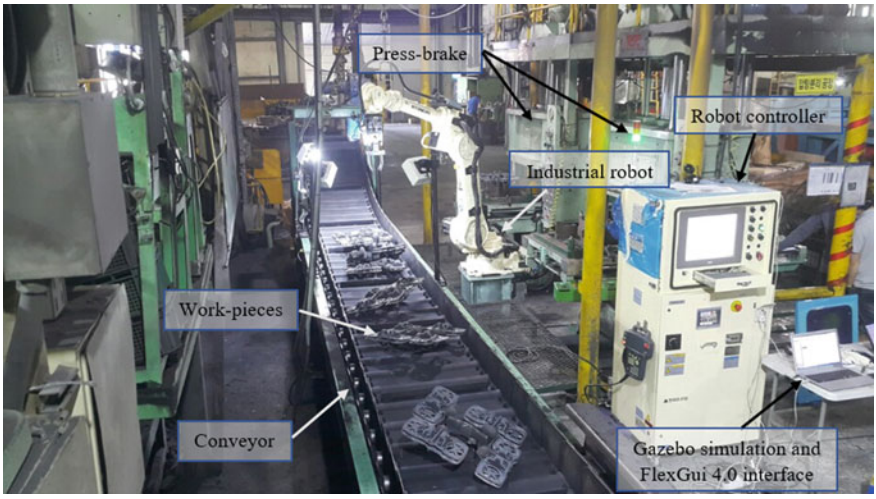


Fig. 6 Industry robot and press-break line in real case

parallel. The conveyor, press-brake, robot controller and all sensors connect to ROS sever, each device will be assigned a node name in ROS. With the help of ROS network, all the objects can easily share information with each other including Gazebo and FlexGui 4.0, achieving Gazebo simulation and FlexGui 4.0 interface control. A demonstration of this communication and supervision for the application can be viewed on the following place: http://t.cn/RSVQGgz.

4 Conclusion

In this paper, we introduce a cyber-physical system, where all sensors, actuators, machines and industrial robot is connected to a local network—ROS network, where they share information easily with each other by ROS message. The novelty of the system that the proposed solution is achieved on the software side, with minimum hardware reconfiguration need. The only hardware configuration is to connect all the present signal to a ROS supported computer, so ROS can access all the devices in the production line. With the aid of FlexGui 4.0, we keep the flexibility of the industrial robot, but extend its' understanding with sensor fusion on a higher decision level, rather than on low robot programming level. Gazebo synchronously simulate production line with the physical world that allows remote monitoring and supervision of the production plant.

Acknowledgements The research was supported by Korea Institute for Advancement of Technology (Kiat) grant funded by the Korea government (N0001698), Eurostars programme (E-9692) and YUJIN Mechatronics Co., Ltd.

References

1. Werr P (2015) IoT evolution world. http://www.iotevolutionworld.com/m2m/articles/401292-how-industry-40-the-internet-things-connected.htm. Accessed 29 May 2017
2. Kagermann H, Wahlster W, Helbig J (2013) Final report of the Industrie 4.0 working group
3. Morgan J (2014) A simple explanation of 'The internet of things'. http://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#3f87e8a66828. Accessed 29 May 2017
4. Harayama Y (2016) Society 5.0. http://fpcj.jp/wp/wp-content/uploads/2016/07/f2d3eec7bf7678840f8adf2ca8000b05.pdf. Accessed 29 May 2017
5. Qvam W (2015) Norge 6.0. https://khrono.no/sites/default/files/walter_qvam_hjernekraft2015.pdf. Accessed 29 May 2017
6. The White House Office of the Press Secretary (2013) President Obama launches advanced manufacturing partnership steering committee "2.0". https://obamawhitehouse.archives.gov/the-press-office/2013/09/26/president-obama-launches-advanced-manufacturing-partnership-steering-com. Accessed 29 May 2017
7. Wübbeke J, Meissner M, Zenglein MJ, Ives J, Conrad B (2016) Made in China 2025: the making of a high-tech superpower and consequences for industrial countries. http://www.iberchina.org/files/2016/MadeinChina_2025_merics.pdf. Accessed 8 Nov 2017
8. Germany Trade & Invest (2016) Industrie 4.0—smart manufacturing for the future. https://www.gtai.de/GTAI/Content/EN/Invest/_SharedDocs/Downloads/GTAI/Brochures/Industries/industrie4.0-smart-manufacturing-for-the-future-en.pdf?v=8. Accessed 29 May 2017
9. Thomas D (2014) ROS/introduction. http://wiki.ros.org/ROS/Introduction. Accessed 29 May 2017
10. PPMAS (2016) www.ppm.no. https://www.ppm.no/flexgui4-Home/Index/downloads. Accessed 29 May 2017

Dynamic Balance System Design and Control for High-Branch Pruning Machine

Yang Li, Shuai Du, Xinxue Zhao, Ziru Niu and Jin Yuan

Abstract The developed high-branch pruning machine is a novel type of high-altitude operation machine which can realize tall tree pruning work. Operators can adjust manipulator posture and cutting position through friendly Human-Machine-Interface (HMI) to control pruning saw realizing pruning operation. Based on the previous research, dynamic balance problem of the machine caused by the posture adjustment of the manipulator is considered in this paper. Firstly, working principle of the high-branch pruning machine, manipulator mechanical structure and basic dynamic balance principle are elaborated. Secondly, the correspondences between manipulator posture and tilting torque, tilting torque and counterweight displacement, are studied based on kinematics analysis and mechanical structure analysis. Thirdly, Consider the numerical calculation and analysis results, the structure of the dynamic balance system and control method are designed. Finally, it is verified by the prototype pruning test in the open area. The test results show that the pruning machine can achieve dynamic balance under the given control algorithm.

Keywords High-branch pruning machine · Dynamic balance · Counterweight control

1 Introduction

Trees-pruning, one major method for forest tending, is beneficial for the trees growth, can improve the trees straightness, bending strength, and wood toughness. It has very positive significance to enhance upper photosynthesis, promote trees growth, and prevent forest fire [1–3]. Mechanized pruning started earlier and widely

Y. Li · S. Du · X. Zhao · Z. Niu · J. Yuan (✉)
Shandong Provincial Key Laboratory of Horticultural Machinery and Equipment, College of Mechanical and Electronic Engineering, Shandong Agricultural University, Tai'an, China
e-mail: jyuan@sdau.edu.cn

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_21

used in Europe and the United States, trees lower than 8 m are mostly pruned by equipment with telescopic arms [4–6]. Manoj Karkee, etc., focused on identification of pruning branches on apple trees in tall spindle architecture, researched branches identification in tall spindle apple trees for automated pruning. Ostraticky firm designed the SL1M pruning machines established for mechanization of pruning especially fruit trees during all the year, which can be used for pruning through the streets or in the municipal services too. The machine prune the ripen branches by the rugged designed pruning bars equipped by set of rotating saw discs. The pruning bars are mounted on the special designed strong suspension made from thick-walled steel profiles. Rinieri present the new orchard pruning machine, which has 400 mm-diameter discs turn at 2700 rpm, can cut wood up to 10 cm of diameter. Xin and Ganhua [7, 8], designed a gasoline engine and an electric motor powered backpack pruning machine, respectively, with gear transmission mode, the operators carry handheld lever to prune branches.

However, the available pruning machineries above mentioned commonly have a narrow cutting range. In order to raise the pruning height, the South China Institute of Tropical Crops Machinery designed and manufactured the 3SG-8 hoist pruning machine, with 8.5 m the maximum lifting height. By manipulating the handle mounted on the work station, the operators control the station position to achieve the pruning work [8]. Jiaoen Zhang, et al. designed a vehicle-mounted high sticks pruning vehicles based on the arms mounted on the end of pruning manipulator, swing mechanism and the auxiliary component to achieve high-altitude pruning operations, assisted by the positioning arm, can increase the working space, reducing the work load. The shortcoming is a special mechanical vehicle is necessary, and having a complex hydraulic system, with high cost of the machine [9, 10].

In order to expand pruning range further, and to improve working efficiency and quality, we designed a novel type of high-branch pruning machinery [11, 12] (Fig. 1a). The dynamic machine balance problem caused by tilting torque in different manipulator postures during pruning mechanical operation is focused on this paper. In Sect. 1, based on the elaboration of basic working principle about counterweight system, the kinematics and mechanical structure analysis is carried

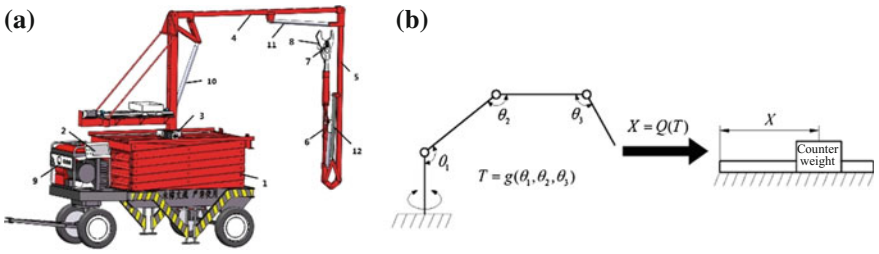


Fig. 1 Pruning machine structure and dynamic counterweight system (*1*—Lift platform; *2*—Human-machine interface; *3*—Turning base; *4*—1st arm; *5*—2nd arm; *6*—3rd arm; *7*—Wireless camera; *8*—Pruning saw; *9*—Power generator; *10*—Electric cylinder 1; *11*—Electric cylinder 2; *12*—Electric cylinder 3.)

out. In Sects. 2 and 3, by research the correspondence between the manipulator posture and the tilting torque, the relationship between tilting torque and the counterweight dynamic displacement are established. In Sects. 4 and 5, the structure design of the dynamic balance system is carried out, and the control method of the dynamic balance of the pruning machine is realized. In Sect. 6, the control algorithm is verified by the prototype pruning test in the open area, and the test results show that under the given control algorithm, the dynamic stability of the machine can be realized.

2 Principle of Dynamic Balance System

The way of dynamic balance system realization is by adjust the counterweight position, produce different rotation torque at the manipulator support point on the rotational base, to counterweigh the torque of the whole pruning machine under different manipulator posture, and stabilize the center of gravity of pruning machine finally.

In the design of dynamic balance system about high-branch pruning machinery, consider that measure machine tilt torque under different manipulator posture by directly sensor installing is high cost and difficult to achieve, we indirectly calculate the real-time tilt torque using electric cylinder telescopic length through the manipulator's kinematics analysis. The actual lengths of the electric cylinders as the input of the balance system, and the counterweight displacement as the output, control the balance of the pruning machine dynamically. The schematic diagram is shown as Fig. 1b, the implementation can be simply described as:

(1) According to the actual elongation of electric cylinders under different posture, derive the rotation angles in the corresponding luffing mechanism between the arms:

$$\theta_i = f(l_i), \quad (i = 1\ ,\ 2\ ,\ 3) \tag{1}$$

where, l_i is the length of electric cylinder i; θ_i is the angle between arm i and the arm $i-1$.

(2) Derive the tilt torque of the pruning machine based on the angles between the arms of the machine under different manipulator postures:

$$T = g(\theta_1, \theta_2, \theta_3) \tag{2}$$

where, T is the tilt torque of manipulator system.

(3) Calculate counterweight displacement by pruning machine tilt torque:

$$X = Q(T) \tag{3}$$

where, X is the theoretical position of the counterweight.

Therefore, the tilting torque of the entire manipulator system is determined based on the amount of elongation of the electric cylinders, and then the counterweight movement can be calculated based on the tilting torque.

3 Manipulator Kinematic Analysis and Tilt Torque

3.1 Kinematic Analysis

The kinematic modeling of manipulator is carried out through D-H method. To facilitate the analysis and calculation, simplify the structure of the pruning machine manipulator as shown in Fig. 2, which totally contains 4 degrees of freedom, respectively, composed of 4 rotating shaft (including the rotary platform).

According to the D-H method to get the mechanical parameters of the manipulator connecting rod as shown in Table 1, where a_i is the length of the connecting rod, α_i is the connecting rod angle, d_i is the rod offset, θ_i is the arm rotation angle.

Further, each arm homogeneous transformation matrix of the pruning machine is obtained in Table 1, as follows:

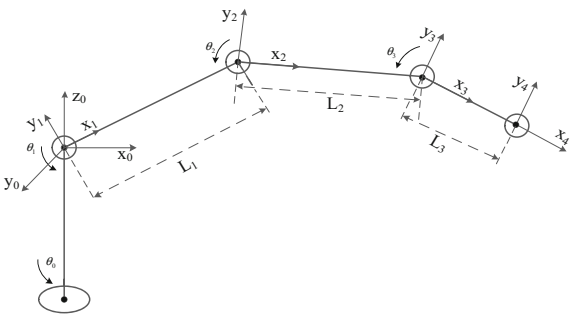


Fig. 2 Manipulator coordinates distribution

Table 1 Manipulator D-H parameters table

i	a_i	α_i	d_i	θ_i
1	0	90°	0	θ_0
2	l_1	0	0	θ_1
3	l_2	0	0	θ_2
4	l_3	0	0	θ_3

$$
{}_{1}^{0}T = \begin{bmatrix} c\theta_0 & 0 & s\theta_0 & 0 \\ s\theta_0 & 0 & -c\theta_0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, {}_{2}^{1}T = \begin{bmatrix} c\theta_1 & -s\theta_1 & 0 & l_1 c\theta_1 \\ s\theta_1 & c\theta_1 & 0 & l_1 s\theta_1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, {}_{3}^{2}T
$$
$$
= \begin{bmatrix} c\theta_2 & -s\theta_2 & 0 & l_2 c\theta_2 \\ s\theta_2 & c\theta_2 & 0 & l_1 s\theta_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, {}_{4}^{3}T = \begin{bmatrix} c\theta_3 & -s\theta_3 & 0 & l_3 c\theta_3 \\ s\theta_3 & c\theta_3 & 0 & l_3 s\theta_3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
$$

where, $s\theta_i$ represents $\sin\theta_i$; $c\theta_i$ represents $\cos\theta_i$.

The kinematics analysis of mechanical pruning manipulator is to calculate end-point position and arm posture, that is homogeneous transformation matrix ${}_{4}^{0}T$, in the case that each joint angle is gotten, where

$$
\begin{aligned}
{}_{4}^{0}T &= {}_{1}^{0}T\,{}_{2}^{1}T\,{}_{3}^{2}T\,{}_{4}^{3}T \\
&= \begin{bmatrix} c\theta_0 & 0 & s\theta_0 & 0 \\ s\theta_0 & 0 & -c\theta_0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} c\theta_1 & -s\theta_1 & 0 & l_1 c\theta_1 \\ s\theta_1 & c\theta_1 & 0 & l_1 s\theta_1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
&\quad \times \begin{bmatrix} c\theta_2 & -s\theta_2 & 0 & l_2 c\theta_2 \\ s\theta_2 & c\theta_2 & 0 & l_1 s\theta_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} c\theta_3 & -s\theta_3 & 0 & l_3 c\theta_3 \\ s\theta_3 & c\theta_3 & 0 & l_3 s\theta_3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\end{aligned} \tag{4}
$$

From matrix calculation, get

$$
{}_{4}^{0}T = \begin{bmatrix} \mathrm{c}\theta_0\mathrm{c}\theta_{123} & -\mathrm{c}\theta_0\mathrm{s}\theta_{123} & \mathrm{s}\theta_0 & \mathrm{c}\theta_0(l_1\mathrm{c}\theta_1 + l_2\mathrm{c}\theta_{12} + l_3\mathrm{c}\theta_{123}) \\ \mathrm{s}\theta_0\mathrm{c}\theta_{123} & -\mathrm{s}\theta_0\mathrm{s}\theta_{123} & -\mathrm{c}\theta_0 & \mathrm{s}\theta_0(l_1\mathrm{c}\theta_1 + l_2\mathrm{c}\theta_{12} + l_3\mathrm{c}\theta_{123}) \\ \mathrm{s}\theta_{123} & \mathrm{c}\theta_{123} & 0 & l_1\mathrm{s}\theta_1 + l_2\mathrm{s}\theta_{12} + l_3\mathrm{s}\theta_{123} \\ 0 & 0 & 0 & 1 \end{bmatrix}
$$

where, $s\theta_{ij}$ represents $\sin(\theta_i + \theta_j)$; $c\theta_{ij}$ represents $\cos(\theta_i + \theta_j)$; $s\theta_{ijk}$ represents $\sin(\theta_i + \theta_j + \theta_k)$; $c\theta_{ijk}$ represents $\cos(\theta_i + \theta_j + \theta_k)$.

3.2 Tilt Torque Calculation

The manipulator posture and position are determinate in a movement through the kinematics analysis, simplify each arm dimension as shown in Fig. 3a, it is easily to get explicit theory function relationship between the tilt torque and the rotation angles based on sketch

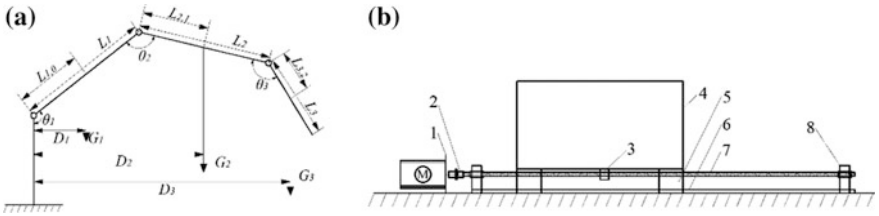


Fig. 3 Manipulator dimension and counterweight system structure (*1*—Motor support; *2*—Coupling *3*—Nut; *4*—Counterweight; *5*—Slider; *6*—Guide rail; *7*—Screw; *8*—Bearing)

$$\begin{aligned} T &= G_1 D_1 + G_2 D_2 + G_3 D_3 \\ &= G_1 L_{1,0} \sin\theta_1 + G_2 \left[L_1 \sin\theta_1 - L_{2,1} \sin(\theta_1 + \theta_2) \right] \\ &\quad + G_3 \left[L_1 \sin\theta_1 - L_{2,1} \sin(\theta_1 + \theta_2) + L_{3,2} \sin(\theta_1 + \theta_2 + \theta_3) \right] \end{aligned} \tag{5}$$

where, L_i is the length of arm i, G_i is simplified arm i's gravity, D_i is the distance between manipulator base and the center of arm i's mass, $L_{i,i-1}$ is the distance between the arm $i-1$'s joint and the center of arm i's mass.

4 Dynamic Balance System Structure Design

4.1 Basic Structural Composition

The main function of the mechanical mechanism of the high-branch pruning mechanical counterweight system is to make the load realize reciprocating motion along the horizontal direction, in order to balance the manipulator system tilting torque. The structure uses a helical drive to convert the rotary motion of the Stepping motor into a linear motion of the counterweight. The system structure diagram is shown in Fig. 3b.

The stepping motor support, the bearing and the linear guide rail are fixed in the corresponding position of the pruning machine respectively. The stepping motor is fixed on its bracket, and the rotation motion of the motor output is transmitted to the screw, by the stepping motor power. The rod drives the nut along the linear motion of the screw shaft and drives the counterweight to move linearly, and the reciprocating motion of the counterweight is realized by the forward and reverse rotation of the stepping motor.

To calculate dynamic counterweight displacement, can simplify the weight of the block diagram shown in Fig. 4, the dynamic counterweight displacement should be:

$$X = \frac{T - m_0 g X_0}{mg} \tag{6}$$

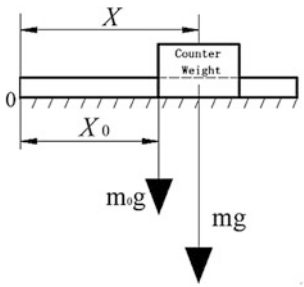


Fig. 4 Dimension diagram of counterweight

where, m is the mass of the counterweight, kg; m_0 is the mass of the counterweight support shelves, kg.

4.2 Control Method of Dynamic Balance System

As mentioned above, under the premise that it is difficult to directly measure the real-time tilting torque of the manipulator system, it is used the length of the electric cylinder as the input value of the balance control system, and the kinematics analysis to calculate the tilting torque by manipulator posture. Finally, the counterweight displacement is taken as the output value, and the dynamic balance control system control flow chart is shown in Fig. 5.

5 Tests and Verification of Dynamic Balance System

In order to determine the total mass of the counterweight system, it is necessary to determine the range of tilting torque that generated by the manipulator system. Import the manipulator system model in SolidWorks to the ADAMS software for pre-processing, and add the drive function for the 3 electric cylinders, the drive function within a fixed time. The amount of elongation is set to the variables DV_1, DV_2, DV_3, and each factor is taken as 5 levels, as shown in Table 2, according to which the parameters of the driving function loaded by the 3 motor cylinders are parameterized. Tilting torque measurements are made on the basement.

Through above method, get the maximum tilting torque of the 125 kinds of manipulators under the different conditions of 3 factors and 5 levels, and the tilting torques can be obtained under each working condition. As the result, the range of the tilting torque is between 424.16 N · m and 1686.40 N · m; the balance arm length is 1 m; the weight of the fixed parts such as the balance arm, the weight

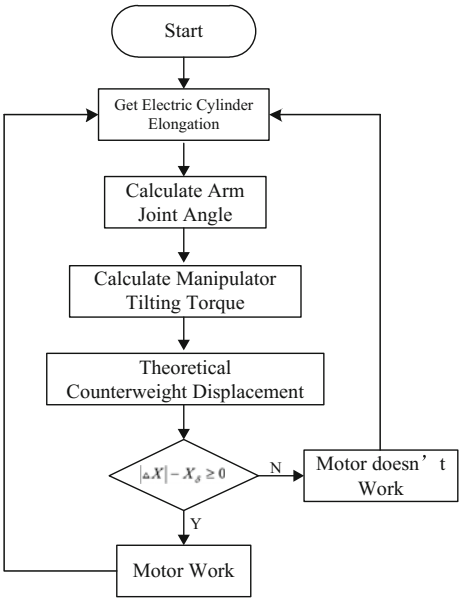


Fig. 5 Dynamic balance control flow

Table 2 Experimental factor parameters

Variables name	Initial value of elongation/m	Maximum value of elongation/m	Level num.
1st Cylinder DV_1	0	0.45	5
2nd Cylinder DV_2	0	0.35	5
3rd Cylinder DV_3	0	0.39	5

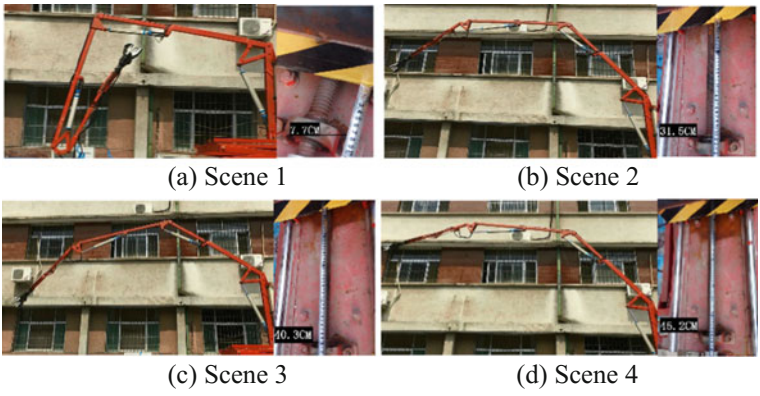


Fig. 6 Test of dynamic balance system

include linear guide, bearing, stepping motor and stepping motor bracket is about 100 kg.

In order to verify the effect of dynamic balance system, take the experimental method as follows: on the horizontal ground with slope of smaller than 5°, various manipulator posture adjustment actions are carried out to realize the arm extension and shrink. And verify the actual distance of the counterweight movement in the current posture, and to verify the effect of the dynamic balance system.

The test results as shown in Fig. 6, with the expansion of the manipulator, under the control of the dynamic balance system, the counterweight is moved in the horizontal direction and played a stable platform.

6 Conclusions

Aiming at the pruning work in forestry, the high-branch pruning machine adapt to 15 m height trees was designed. Based on the previous design, the dynamic manipulator balance problem is researched in this paper, and the following work is finished:

(1) The working principle of the dynamic balance system is introduced. According to the mechanical structure of the high-branch pruning machine, the manipulator kinematics is analyzed, and the relationship between the manipulator posture and the tilting torque is established.
(2) The dynamic balance system structure design is carried out, and the basic structure of the dynamic balance system is established, and the control method of the dynamic leveling system is studied. The controller and the control method are applied to the prototype test. The experimental results show that under the action of the dynamic balance control method, the counterweight automatically adjusts the position according to the attitude of the manipulator, achieved the effect of balancing the tilting torque. The valuable experience for the further development of the high pruning machine is accumulated.

References

1. Zhang X, Cao B, Zhang X et al (2010) Progress in research on theory and technology of poplar pruning. Hebei J For Orchard Res 25(3):235–239
2. Chunsheng W, Zhigang Z, Longdun W et al (2012) Effects of pruning height on trunk borer (Arbela spp.) damage to betula alnoides. J Northwest For Univ 27(6):120–123
3. Siyi D, Jiyun S, Qingpeng Y et al (2015) Effects of thinning and pruning on soil microbial biomass carbon and soil enzyme activities in Chinese fir plantation. J Cent S Univ For Technol 35(6):75–79
4. Zhengping G, Ruizhen S (2004) Present status and prospect of gardening machines at home and abroad. For Mach Woodwork Equip 32(2):4–7

5. Qiming J, Xiujuan M, Dahai Z (2001) Discussion on the mechanization of pruning tall trees in cities. J Liaoning For Sci Technol 3:43–44
6. Shiwei L, Shuyang W, Hui W et al (2008) Present status and prospect of tree pruning machinery. For Mach Woodwork Equip 36(1):15–16
7. Xin J, Tang X, Quan L et al (2003) On compact adjustable tree pruner. J Hunan Agric Univ (Nat Sci) 29(1):63–64
8. Qianhua Y, Beiping X (2010) Structural design and research on electrically operated hedgerow pruner. For Sci Technol 35(1):48–50
9. Enzhang J, Jianhong Z (1990) A study of lorry-mounted highlift pruner. J Nanjing For Univ 14(1):63–67
10. Kui L, Jinbao L, Jing Z et al (2014) The development situation and tendency of grape vine pruning machine. J Agric Mech Res 3:246–248
11. Chen Y (2016) Design and optimization analysis of tall trees pruning machine. Shandong Agricultural University
12. Yuan J, Wu J, Cao B et al (2013) One of poplar automatic pruning device. Chinese Patent, ZL201310046491.9, 2013

Braking Distance Monitoring System for Escalator

Xiaomei/XM Jiang, Shuguang/SG Niu and Lanzhong/LZ Guo

Abstract In the design, manufacture, use and test of escalator, the measurement of braking performance is very important. The braking distance of escalator is the most important parameter to judge its braking performance. In this thesis an effective mode to use wireless remote control technology as information transport platform for wireless monitoring system, with advantages of simple principle, high security, wide spread and so on, its hardware structure and software module are described, achieving dynamic display of data, analysis, remote control and data management, valuing important in practical use.

Keywords Escalator · Braking distance · Wireless trigger device SCM

1 Introduction

As a kind of special equipment for escalator to carry personnel, its safety performance is very important. Braking distance of escalator is an important measure of braking performance which will directly determine the safety of the passengers. So the braking performance should meet the braking requirements, so that the escalator can stop at the prescribed braking distance and deceleration speed [1, 2]. Starting from the design requirements of escalator braking system, Escalator braking system should be able to make the escalator have a braking process near uniform deceleration to slow down until stopping, and keep it stop (brake). The braking system is based on friction torque to brake the drive main motor, its mechanics actually is by friction drag between the brake pulley and brake shoe, making all moving parts

Xiaomei/XMJiang (✉) · Shuguang/SGNiu · Lanzhong/LZGuo
School of Mechanical Engineering, Changshu Institute of Technology, Changshu, China
e-mail: jszjjxm@hotmail.com

Xiaomei/XMJiang · Shuguang/SGNiu · Lanzhong/LZGuo
Jiangsu Key Construction Laboratory of Elevator Intelligent Safety, Changshu, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_22

including the carrying capacity on the stair step dissipate the inertia energy and stop its movement within the specified distance which is the braking distance [3].

If the escalator braking distance does not meet the design requirements, the possible harm will be very large. Although before the measure is simple and fast, but it also has obvious shortcomings, such as poor measurement accuracy, the strong influence of the subjective judgment of the operator, which can not meet the needs of the current society [4]. Therefore, it is very necessary to research and develop a kind of advanced and practical testing equipment for escalator. Therefore, to obtain braking distance data conveniently, accurately, objectively is a very important job. We want to develop an automatic tester which can realize the automatic, fast and accurate measurement of the braking distance of the escalator. The General Administration of quality supervision inspection shall, in accordance with the rules for the supervision and inspection of escalators, regulate the methods for measuring the braking distance of escalators [5]. The braking distance should be measured when the electrical stop device is operated:

1. Brake distance check without load;

   (1) Make markers on the step, pedal or adhesive tape and apron board;
   (2) Cut off the power supply when operating escalator or automatic sidewalk to align markers coincide;
   (3) Measure the distance between the two markers to meet the requirements.

2. Brake distance check with load

   (1) Determine the braking load:

      a. step load
         Step width 0.6 m, 60 kg; step width 0.8 m, 90 kg; step width 1.0 m, 120 kg;
      b. brake load = step load * lifting height/maximum visible step riser height

3. Distribute the total braking load in the upper 2/3 steps of escalator, start down the escalator, immediately cut off the power supply once in normal operation, and check whether the braking distance is consistent with the requirements.

Although this method is simple and fast, it also has the obvious shortcomings which cannot meet the needs of the current society. Therefore, it is very necessary to research and develop a kind of advanced and practical testing equipment for escalator. Through the optimization to improve the performance of the entire monitoring system will be a promising also great topic, take this exploration in the future will wide spread in escalators.

2 Design Principle and Requirement

From the kinematic formula (1), the stopping distance depends on the braking deceleration, and the braking deceleration can be calculated by the formula (2) [6, 7].

$$L = v^2/2a \tag{1}$$

L: stopping distance (m); v: running velocity of stair step (m/s); a: braking deceleration (m/s^2)

$$a = 2\varepsilon n_z R_t/\pi n_t \tag{2}$$

ε: angular deceleration (rad/s^2); n_z: brake wheel rotation speed (r/min); n_t: step sprocket rotation speed (r/min); R_t: pitch radius of step sprocket (m)

$$M_g = QgR_t n_t \sin\alpha/n_z \tag{3}$$

Q: the mass of the step load (kg); g: gravity acceleration (m/s^2); α: Tilt angle of escalator (°).

According to the requirement of GB16899-2011, the braking torque M_z is assumed only considering the influence of the gravity additional torque M_g when going down with load, and the actual angular velocity of the brake wheel can be calculated:

With load

$$\varepsilon = (M_Z - M_g)/J_{ZQ} + J_{ZK} \tag{4}$$

Without load

$$\varepsilon = M_Z/J_{ZK} \tag{5}$$

M_Z: braking torque (Nm); M_g: additional torque of gravity (Nm); J_{ZQ}: rotary inertia conversion of the load to the brake wheel axle (kgm^2); J_{ZK}: rotary inertia conversion of the step rotating parts and motor to the brake wheel axle without load (kgm^2).

From the above formula the following can be obtained:

The stopping distance of escalator when with load:

$$L = \frac{\pi v^2 n_t}{4n_z R_t} \times \frac{J_{ZQ} + J_{ZK}}{M_Z - M_g} \tag{6}$$

The stopping distance of escalator when without load:

$$L = \frac{\pi v^2 n_t}{4n_z R_t} \times \frac{J_{ZK}}{M_Z} \tag{7}$$

It can be seen that for the same brake torque of the same escalator, the main factors that affect the braking distance are the mass of the load. With the increase of load, the stopping distance will be increased when going down with load. If

ignoring the influence of the gravity additional torque M_g on the braking torque, the increase of the braking distance is proportional to the increase of the load on the stair step.

In GB 16899-2011, there are two sets of safety devices related to braking distance. If the braking distance is too large, it will make the two safety devices act, cannot effectively play a protective role. At the same time, the braking distance is too large, then the braking torque is too small which cannot stop reliably, especially in the event of a reversal, there will be a risk prone to stampede [8]. The stopping distance is not as small as possible. For escalator nominal velocity 0.5 m/s, if you want to ensure braking deceleration ≤ 1 m/s^2, obtain the minimum system stopping distance 0.125 m according to formula (1). If the stopping distance is too small, it will make deceleration increase which is easy to cause the passengers on the escalator to fall down, serious occasion will also result in a stampede.

3 Braking Distance Testing Techniques

The measurement uses the principle of the encoder, advanced microcontroller technology and wireless remote control technology to control the measure. The monitoring system is composed of a host machine and a wireless trigger device. The host machine is equipped with a single chip microprocessor, a keyboard, a power supply, a wireless receiving chip and a rotary encoder connected with the single chip microprocessor through a signal conversion circuit, and a driving wheel which is coaxially arranged with the rotary coding wheel in the rotary encoder and is driven by the skirt part of the escalator etc., and a wireless transmitting chip, a battery and a linkage button switch are arranged in the wireless trigger device which is connected with the emergency stop button arranged on the main circuit of the escalator.

The host machine is placed at the junction of the comb plate and step board of escalator, and wireless trigger device placed on the stop button on the escalator, when the testing operators let the escalator start running at rated speed, then judge the escalator in stable running state, choose the appropriate time to quickly press the button switch on wireless trigger device of the escalator, the escalator immediately stop running, the monitoring system can calculate the braking distance of escalator according to the measured encoding pulse signal. Schematic diagram of installation position is shown in Fig. 1.

The monitoring system composition is shown in Fig. 2. The host machine is composed of a rotary encoder, a signal conversion circuit, a single chip microcomputer, an infrared transmitting and receiving device, a liquid crystal display screen and a micro printer. The driving wheel coaxial mounting with the rotary encoder is in direct contact with the stair step by lifting device adjustment, and when the escalator is running, the driving wheel is driven by stair step to rotate simultaneously. When the wireless trigger device is pressed, the wireless signal is sent out, and the encoder starts to count the number of pulses until the elevator

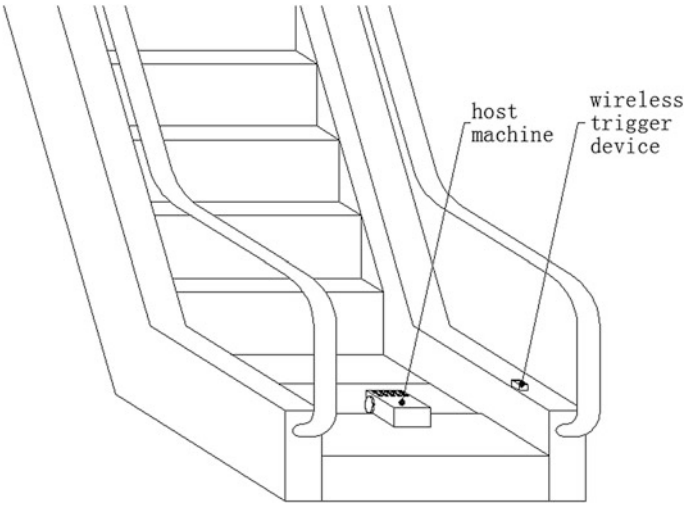


Fig. 1 Schematic diagram of installation position

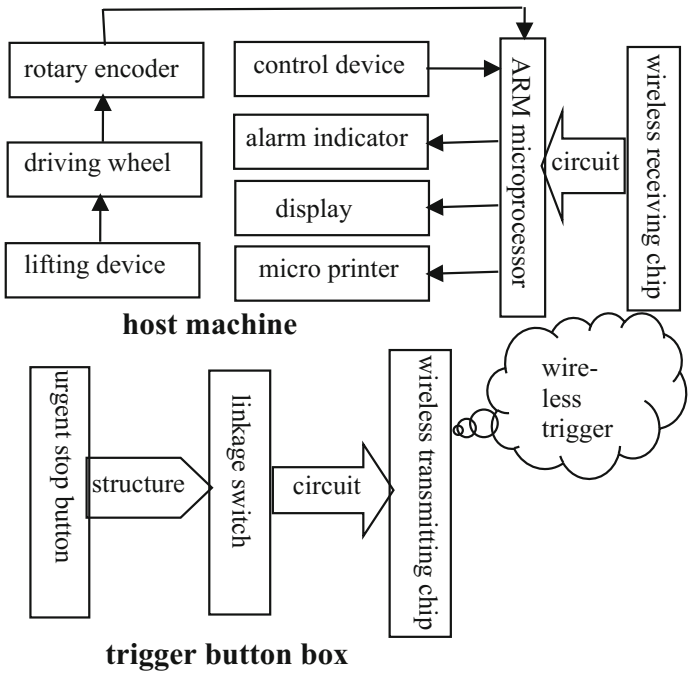


Fig. 2 Components of monitoring system

stops. The rotary encoder is installed on a flexible shaft with a large degree of freedom, so as to be convenient for testing and adjusting.

Because each pulse is corresponding to certain angle of rotation of the encoder, when wireless touch button is pressed, timer starts until the escalator stops, the encoder turned a certain angle, angle conversion to radian; the product between radian and the radius of the wheel flange is the stopping distance. After data has been processed by the microcontroller, displayed directly on the screen and printed by micro printer. The calculation principle is as follows, the escalator braking distance is L, round wheel radius of the rotary wheel is R, a circle revolution of round wheel need n_0 pulses (12-bit encoder, $n_0 = 4096$), that is, one revolution of the encoder requires 4096 pulses, each pulse of the encoder, the rotation angle of only about 0.09°. The escalator need total n pulses from main circuit power off to stop, the escalator braking distance L is:

$$L = n/n_0 \times 2\pi R \tag{8}$$

The system can obtain encoding data through the given time period (sampling period) to calculate the escalator running velocity, maximum braking acceleration and average braking acceleration after braking, the principle of these data acquisition is shown in Fig. 3. When the escalator is running, the monitoring system in the testing standby state, when the elevator emergency stop button is pressed, the linkage switch is trigged to control infrared remote transmitting chip to transmit signal, the host machine receives the trigger signal of infrared remote control, CPU control rotary encoder to start counting [9, 10]. After the counting results are sent to the CPU, which stored preparation program, according to the keyboard set function, corresponding maximum braking acceleration, braking distance, braking speed, average acceleration can be calculated. CPU sends the data through the serial port for the micro printer to print. Main program flow block diagram are shown in Fig. 3.

To obtain the starting point of calculating the number of pulses is an important work of this design. This design uses PT2262 and PT2272 to make wireless trigger receiving device. The transmitting chip PT2262-IR integrates the carrier oscillator, the encoder and the transmitting unit, which makes the transmitting circuit very simple. coded signal which is sent by encoding chips PT2262 comprises the address code, the data code, synchronization code to form a complete code, after decoding chip PT2272 receiving the signal, the address code checking twice, then VT pin outputting high level, at the same time, the corresponding Data pin high output, if the sending terminal has been holding down the key, encoding chip will continuously transmit. When the transmitter does not press the button, the PT2262 is not connected to the power supply, the 17 pin is low level, so the high frequency transmitting circuit does not work; when a button is pressed, PT2262 work electrically, the pin 17 outputs serial data signal through the modulation, when the 17 pin in high level, high frequency transmission circuit oscillate and transmit the constant amplitude high frequency signal, when the 17 pin in low level, the high

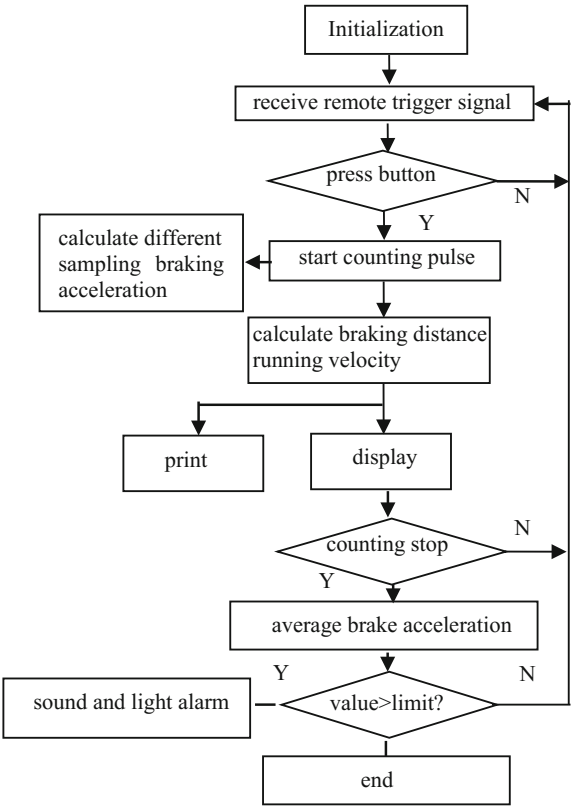


Fig. 3 Main program flow block diagram

frequency transmitting circuit stop oscillation, Therefore, the high frequency transmitter circuit is completely controlled by the digital signal output of 17 pin of PT2262, thus the amplitude shift keying (ASK modulation) of the high frequency circuit is equivalent to amplitude modulation of modulation 100%.

The system used the principle of infrared transmitting and receiving remote control, diagram of infrared emission circuit composed of PT2262 is shown in Fig. 4 and infrared receiving circuit composed of PT2272 is shown in Fig. 5. In order to demodulate the modulated encoding signal correctly, the receiver adds a pre amplifier stage to ensure that the signal amplitude of the input to PT2272 is large enough. The control signal is connected to the I/O interface of SCM through the VT interface of PT2272 output.

The braking distance measurement based on ARM embedded controlling system is used in the monitoring system of escalator, in the process at each stop, accurately measure the length of the braking distance, if the value is beyond the set limit, alarm and fault code is displayed, the escalator cannot be started again, effectively guarantee the braking force of the brake to safeguard passenger safety.

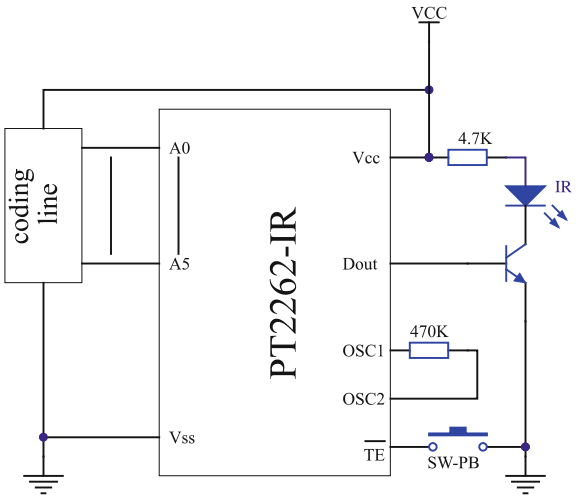


Fig. 4 Infrared transmitting circuit

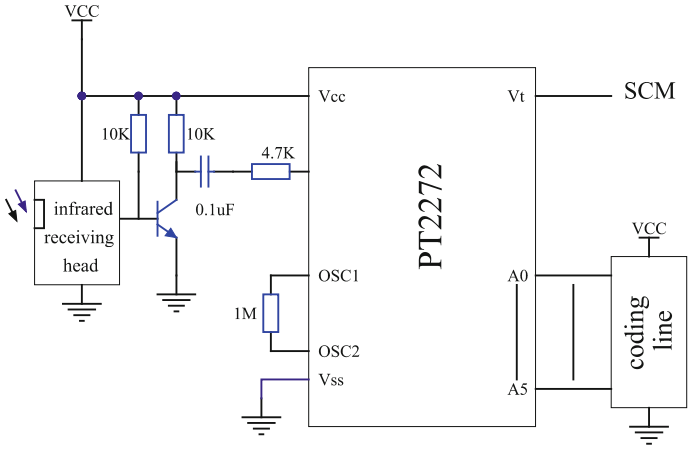


Fig. 5 Infrared receiving circuit

4 Conclusion

The traditional measurement of braking distance using wired mode mainly through the clamp current sensor to collect the signal of power off, triggering the host machine to start pulse signal calculation, and also dismantling the cover of the drive station or corner station, the cover cannot be restored in original position when measurement, mixing circuit lines and moving parts together will produce a certain security risk, and using wireless transmitting and receiving device, no need to

remove the cover, and will not generate the risk of electrical lines contact with moving parts. At the same time, the system adopt wireless remote control technology, can facilitate the layout installation in test site, use more conveniently. Because the system sets up the host machine and trigger button box separately, the use of wireless remote control commands to control, the innovative structure and technical methods ensure that the test system has a convenient field test operation, the appropriate design can be used for escalator, automatic sidewalk, elevator and crane to measure running speed and running distance, braking distance, braking acceleration.

This monitoring system can replace the backward manual measurement method which has been used for a long time, obtain objective and accurate measurement data, can provide more objective and accurate data for the user and the testing organization, using the wireless trigger device is easy for the layout of the test site, it is more convenient to use with high accuracy, simple operation and convenient installation, ensure the convenient and superior performance of on-site testing operation and can be applied to a variety of test occasions.

Acknowledgments This work is supported by Natural Science Research Major Project of Higher Education Institution of Jiangsu Province (No. 17KJA460001).

References

1. GB 16899-2011 the manufacture and installation safety norm of escalator and automatic sidewalk [S]
2. GB 7588-2003 elevator manufacture and installation safety norm [S]
3. Wang W, Lu J (2006) Calculation of braking ability of escalator and automatic sidewalk [J]. Mech Des 2006(5):18–19
4. GB/T 31821-15 scrap technical conditions for main components of elevator [S]
5. Rules for the supervision, inspection and regular inspection of elevator—escalator and automatic sidewalk, the State Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China (2012) [Z]
6. Lianggui P (2006) Mechanical design [M], 8th edn. Higher Education Press, Beijing
7. He S (2007) Engineering mechanics [M]. Shanghai Jiao Tong University Press, Shanghai
8. Wang G, Liu S (2007) Elevator braking distance and detection device [J]. Elevat Install 2007 (12): 23–24
9. Jiang X (2015) Research on vibration control of traction elevator. In: International industrial informatics and computer engineering conference (IIICEC 2015), Xian, pp 2144–2147
10. Jiang X, Hua Z, Rui Y (2013) Research on intelligent elevator control system. Adv Mater Res 605(11):1802–1805

Design of Large Tonnage Elevator Sheave Block

Lanzhong Guo, Xiaomei Jiang and Yannian Rui

Abstract Large tonnage elevator lifting heavy cargos have been more and more widely used. In order to ensure the normal operation, maximize the function of a cargo elevator, avoid safety accidents, improve the use efficiency, achieve cargo elevator standardization, optimization design is more important on the traction sheave block structure of large tonnage cargo elevator, we seek to find a breakthrough from the common traction sheave block structure and accordingly the sheave structure can own a better performance in the application of large tonnage traction elevator. In this paper with sheave block rated load 10,000 kg, the rotary speed 60 rpm, pitch diameter 640 mm, equipped with wire rope 8 strands Φ16, rope groove spacing 25 mm, the parameters of the sheave block are given to design the sheave structure suitable for large tonnage elevator for load capacity 15–20 ton.

Keywords Large tonnage elevator · Sheave block · Structure design
Static strength check

1 Introduction

In 1903, Otis Elevator Company transformed from drum driven elevator to traction drive and built the foundations for the modern long stroke elevator sheave and wire rope in traction drive system is an important loading component in the elevator which carries all the mass load of the elevator. In recent years, this kind of special equipment for elevator, especially large tonnage cargo elevator, relevant content published in various scientific literature library collections and related professional

L. Guo · X. Jiang
Jiangsu Key Construction Laboratory of Elevator Intelligent Safety, Suzhou, China

X. Jiang (✉)
School of Mechanical Engineering, Changshu Institute of Technology, Changshu, China
e-mail: jszjjxm@hotmail.com

Y. Rui
College of Mechatronics Engineering, Soochow University, Suzhou, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_23

journals can be found, but there is almost no discussion for sheave block itself about large tonnage traction elevator, the study will commence the sheave block structure design less attention by scholars, seek to find out the breakthrough from the traction sheave structure and use in large tonnage traction elevator, design sheave block structure with good performance under the premise of ensuring safety. The «GB 7588-2003 elevator manufacture and installation safety norm» as well as the «GB 25,856-2010 freight elevator manufacture and installation safety norm» are referred to give dual attention to parameter requirements that the non-specialized operation security and equipment itself needs to satisfy on large tonnage elevator [1, 2]. Through the optimization of the structure to improve the strengthening performance of the entire mechanical system will be a promising also great topic, take this exploration in the future will spread in large tonnage traction elevator [3].

2 Structure Design of Sheave Block

2.1 Sheave Traction Ratio and Rope Winding Selection

The sheave traction ratio which is also called rope ratio refers to the circumferential speed of the traction sheave namely the ratio of rope moving speed to the car vertical lifting speed, when ignoring the movable sheave weight and the sliding friction between wire rope and sheave, friction between the sheave internal structure in the ideal case, the input and output power should be the same [4], so the lifting weight is far greater than the weight of the movable sheave on large tonnage elevator, laborsaving ratio can be approximately regarded as being inversely proportion to traction ratio [5].

Considering mainly the design of large tonnage traction elevator, sheave block traction ratio are given priority to labor-saving effect, by increasing the ratio as far as possible to make sheave block with each sheave 10,000 kg load obtain greater rated traction load.

From the basic properties of the sheave, in order to make the sheave block energy-conserving, namely under the premise of each sheave rated load 10,000 kg, the final greater payload are calculated, sheave block traction ratio required should be larger. The total quality of all the movable sheave components will also be included in the total load, the final calculated payload affected, under the condition that rated load and rated speed and sheave pitch diameter are consistent, when each movable sheave increase, the additional non payload value is a certain value, at this time the smaller traction ratio can reduce the number of movable sheave, the proportion of payload in the total load be increased [6].

How energy-conserving is not only related to the number of movable sheave in the sheave block, but also affected by the winding method of the steel wire rope in the sheave block. At the same time, because each sheave is equipped with steel wire rope 8 strands Φ16 in a large tonnage traction elevator, so the rope winding method

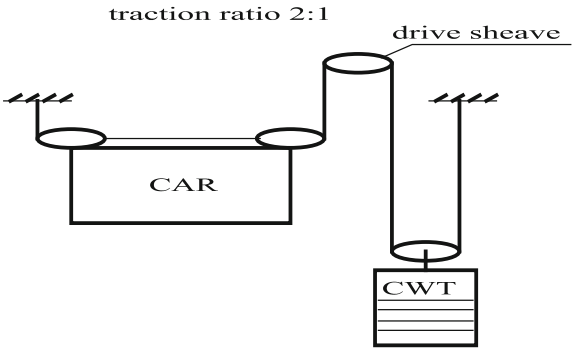


Fig. 1 Sheave rope winding method with traction ratio 2:1

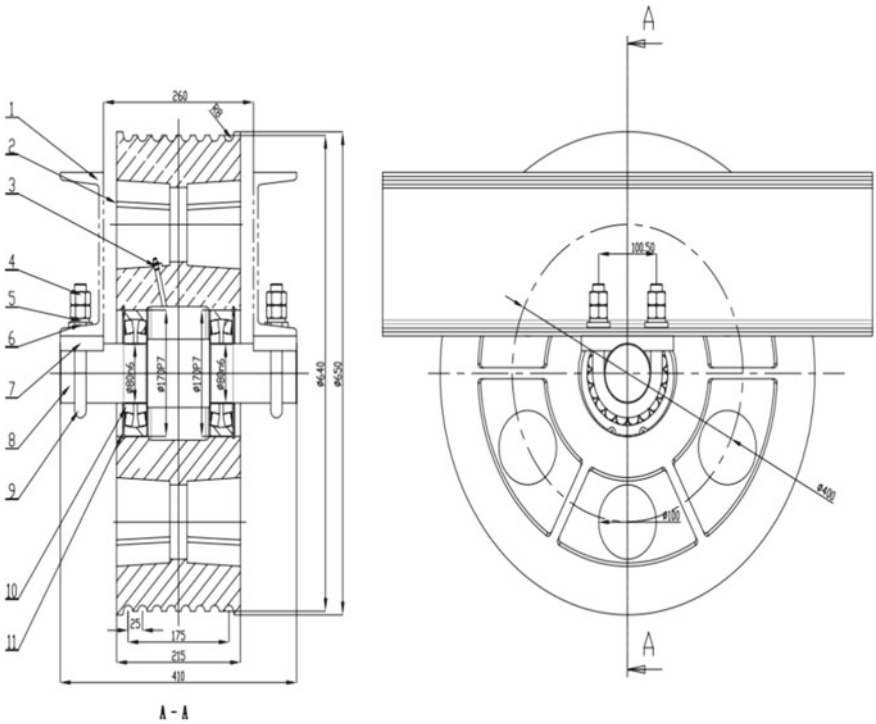
should be considered with different segments of wire rope separately running over the sheave placed in the same vertical plane as far as possible to simplify the wire rope and the relevant contact stress, and avoid excessive lateral force damaging rope groove of the sheave [7]. In addition, the winding method of the steel wire rope is closely related to the elevator machine room type. After comprehensive consideration, select the sheave rope winding method with traction ratio 2:1 is shown in Fig. 1.

2.2 Design of Basic Structure of Car Top Sheave Block

In order to facilitate the manufacture and assembly, the sheave structure in the sheave block on car top and the sheave structure in the sheave block on counterweight top are not separately designed. Because the car top sheave shaft and the counterweight top sheave shaft have rated load 10,000 kg, rated speed 60 rpm and the load type is mainly the radial load, select bearing GB288-1994 21316C [8]. The basic structure of the sheave is shown in Fig. 2.

Because a car top sheave and car connection is not the same as a counterweight top sheave and counterweight connection, the different connection structure design for a car top sheave and counterweight top sheave. In the structure sheave shaft8 placed on the car top connecting the car top mainly hoop U-bolts9 at both sides of sheave shaft and channel steel hanger plate1 connected to the car frame.

The sheave shaft in sheave block at the car top, mainly carry the radial load transferred from the bearing inner ring, rated load 10,000 kg and gravity of related parts in sheave block at car top, for ease of calculation and analysis, ignoring the gravity effect of all relevant parts, only considering the radial load from the bearing inner ring, simplify the action point to axial center of the bearing, the rated load 10,000 kg is equally distributed to two bearings, each carries 5000 kg, under the condition of gravity acceleration g = 9.81 m/s^2, each bearing only carries the radial load 49,050 N, the constraints of the sheave shaft in sheave block on car top are placed mainly through hooping U-bolts at both sides of the sheave shaft and



1. hanger plate 2.sheave 3. Pressure oil filling cup 4. Nut M20
5. spring pad 6. square taper washer for U-steel 7. base plate
8.sheave shaft 9. U-bolt 10.bearing 11. circlip

Fig. 2 Basic structure of sheave on car top

connecting to the car frame by channel steel hanger plate, the constraint point is simplified as a contact point of sheave shaft with U-bolt, the load analysis of sheave shaft in sheave block at the car top is simplified, and the shear stress diagram and bending moment diagram are shown in Fig. 3.

The 21316C bearing inner diameter 80 mm, thus simplifying the calculation diameter of the sheave shaft on the car top ϕ 80 mm, from the shear stress diagram, maximum shear force of the sheave shaft F_{smax} = 49,050 N, calculating the maximum shear stress:

$$\tau_{1\max} = (4F_{s\max})/(3\pi R^2) = 13.01 \text{ MPa} \quad (1)$$

in the bending moment diagram, the maximum bending moment of the sheave shaft M_{max} = 4659.75 N*m, calculating maximum bending stress:

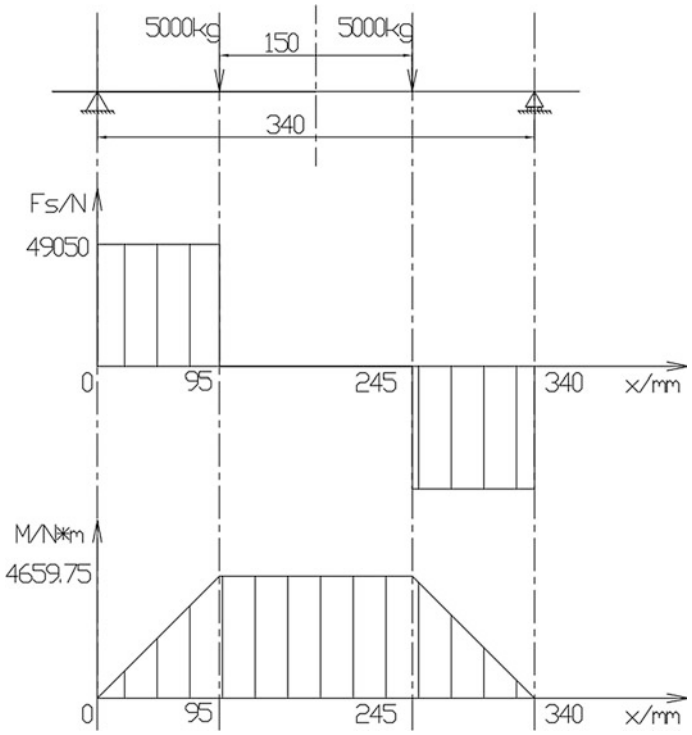
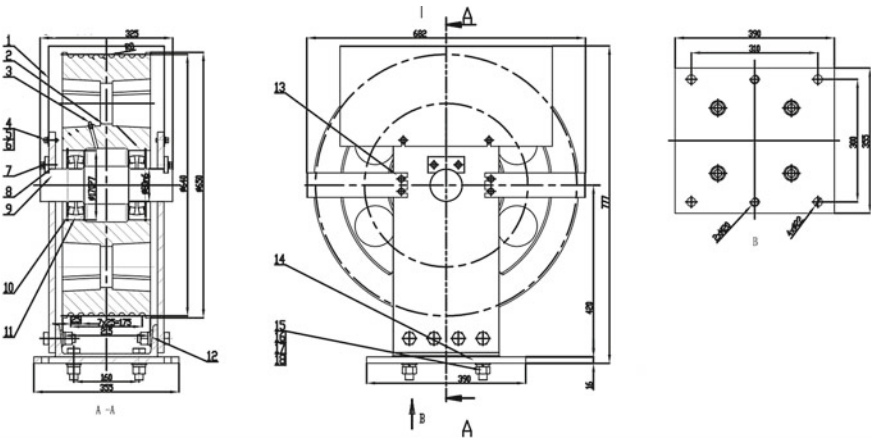


Fig. 3 Mechanical analysis of car top sheave shaft

$$\sigma_{1\max} = (M_{\max)}/(W) = (32M_{\max})/\pi d^3 = 92.7 \text{ MPa} \quad (2)$$

2.3 *Design of Basic Structure of Counterweight Top Sheave Block*

For counterweight top sheave block the structure is shown in Fig. 4. In the connecting structure the counterweight sheave shaft9 at counterweight top connect the counterweight top by stuck counterweight sheave shaft clamp7 on both sides of counterweight sheave, counterweight sheave side plate8 and connected to counterweight sheave base14 on the counterweight frame. The constraints of the sheave shaft in sheave block on counterweight top are placed mainly through stuck counterweight sheave shaft clamp on both sides of counterweight sheave, counterweight sheave side plate8 and connected to counterweight sheave base14 on the counterweight frame, the constraint point is simplified as axial center of the contact part between the sheave shaft and the counterweight sheave shaft clamp.



1. counterweight sheave cover 2. sheave 3. Pressure oil filling cup 4. Bolt M8×25
5. spring pad 8 6. Flat washer8 7. counterweight sheave shaft clamp
8. Counterweight sheave side plate 9. Counterweight sheave shaft 10. bearing
11. circlip 12.Counterweight sheave side cover 13.Counterweight sheave base
14. Bolt M8×25 15. Flat washer20 16. spring pad20 17. Nut M20

Fig. 4 Basic structure of sheave on counterweight top

The load analysis of sheave shaft in sheave block at the counterweight top is simplified, and the shear stress diagram and bending moment diagram are shown in Fig. 5.

Because the cross section of the contact part between sheave shaft and counterweight sheave shaft clamp in sheave block on the counterweight top is not a complete circle, thus simplifying calculation diameter of the sheave shaft on the counterweight top ϕ 70 mm, from the shear stress diagram, maximum shear force of the sheave shaft on counterweight top F_{smax} = 49,050 N, calculating the maximum shear stress:

$$\tau_{2\max} = (4F_{s\max})/(3\pi R^2) = 16.99 \text{ MPa} \tag{3}$$

In the bending moment diagram, the maximum bending moment of the sheave shaft M_{max} = 3715.54 N*m, calculating maximum bending stress:

$$\sigma_{2\max} = (M_{\max})/(W) = (32M_{\max})/\pi d^3 = 110.39 \text{ MPa} \tag{4}$$

2.4 Life and Static Strength Check

Basic rated dynamic load of bearing 21316C C = 355,000 N, ignoring gravity effect of related parts in sheave block, only considering the radial load transferred

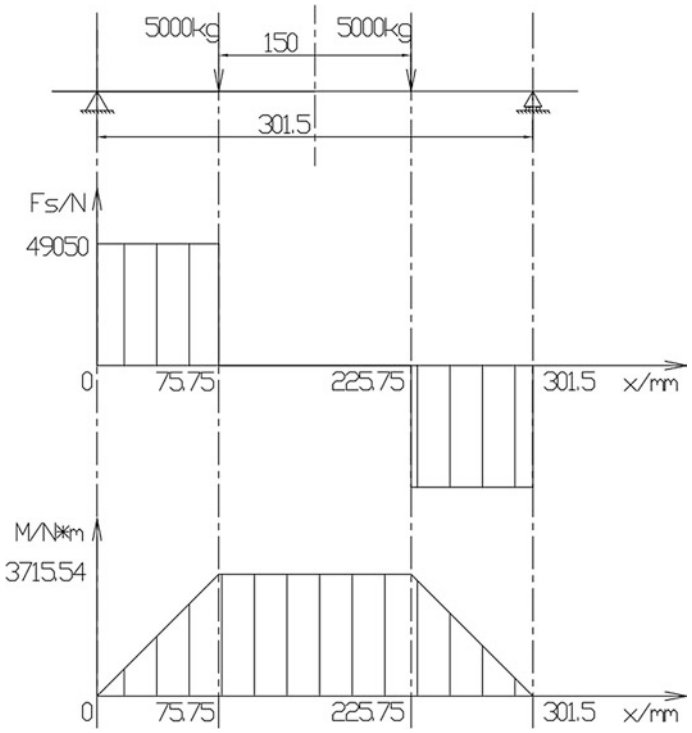


Fig. 5 Mechanical analysis of counterweight top sheave shaft

from the sheave to bearing outer ring, namely the rated load of sheave, because two bearings are completely same, the rated load 10,000 kg is equally distributed to two bearings, each carries 5000 kg, under the condition of gravity acceleration g = 9.81 m/s^2, each bearing only carries the radial load 49,050 N, regards it as the equivalent dynamic load, F = 49,050 N.

From the bearing service life formula,

$$L_{10h} = [(10^6)/(60n)] \times [(f_t C)/f_f F)]^{\varepsilon} \tag{5}$$

in which take temperature coefficient $f_t = 1$ (operating temperature lower than 105 °C), load factor $f_f = 1.2$ (slight impact), life index $\varepsilon = 10/3$ (in roller bearing), and basic dynamic load rating C = 355,000 N, rated speed n = 60 rpm, the equivalent dynamic load F = 49,050 N, finally obtain service life $L_{10h} =$ 11,0932.25 h through the calculation, according to daily usage time of revolution is 18 h, this bearing may be used approximately 6163 days, namely service life of this bearing is roughly above 16 years in the normal service condition, so for bearing GB288-1994 21316C, the service life check result is qualified.

Basic rated static load of bearing 21316C C_0 = 355,000 N, according to the same calculation method above, each bearing only carries the radial load 49,050 N, regards it as the equivalent static load, F_0 = 49,050 N, take static strength safety factor S_0 = 1.5 (high requirement for rotation accuracy and smooth operation in rotary bearing), calculation formula F_0S_0 = 73575 N, compared with the basic static load rating C_0 = 375,000 N, $C_0 > F_0S_0$, so for bearing GB288-1994 21316C, static strength check for qualified.

A car top sheave shaft and a counterweight top sheave shaft select the same material 45 steel, its yield limit $\tau_s \geq 355$ MPa, for plastic material 45 steel, allowable stress equal to yield limit divided by safety factor S_0 (1.5–2), where S_0 = 2, then calculate the allowable bending stress $[\sigma] = \sigma_s/S_0 = 177.5$ MPa; at the same time the allowable shear stress is 0.5–0.6 times allowable stress, proportional coefficient k = 0.5, then calculate the allowable shear stress $[\tau] = [\sigma]/k = 88.7$ MPa.

By the mechanical analysis calculation result of sheave shaft in sheave block at the car top, the maximum shear stress $\tau_{1max} = 13.01$ MPa, the maximum bending stress $\sigma_{1max} = 92.7$ MPa, by comparison, $\tau_{1max} \langle [\tau]$, $\sigma_{1max} \langle [\sigma]$, so for car top sheave shaft the static strength check is qualified.

By the mechanical analysis calculation result of sheave shaft in sheave block at the counterweight top, the maximum shear stress $\tau_{2max} = 16.99$ MPa, the maximum bending stress $\sigma_{2max} = 110.39$ MPa, by comparison, $\tau_{2max} \langle [\tau]$, $\sigma_{2max} \langle [\sigma]$, so for counterweight top sheave shaft the static strength check is qualified.

3 Finite Element Analysis

SOLIDWORKS Simulation can calculate displacement, strain and stress of parts under the interior and exterior load using finite element method displacement formula. Through modeling bearing and sheave shaft in the sheave block at the car top and the counterweight top in SolidWorks, carry on finite element analysis using built-in SolidWorks Simulation under the following conditions: linear material, small deformation (structural impact), static load.

The comparison design and analysis are carried out and the rolling bearing 21316C GB 288-94 model is established in SolidWorks software, and solid automatic mesh generation of the model was done. The material of the rolling bearing is defined as AISI 1020, the elastic modulus is 200 GPa, the yield stress is 3.516 GPa, and the Poisson's ratio is 0.29 [9]. Applying load and boundary conditions, the equal and opposite load 18 ton is placed on both sides of the bearing the static stress analysis results is as follows. Figure 6 is the stress contour under test load, from the graph we can see that the maximum stress of rolling bearing is 18.1 MPa, far less than the yield stress of the material, which can better meet the strength requirements of the large tonnage elevator. Static stress analysis result for maximum resultant displacement 0.007 mm, the maximum strain 7.5e−5.

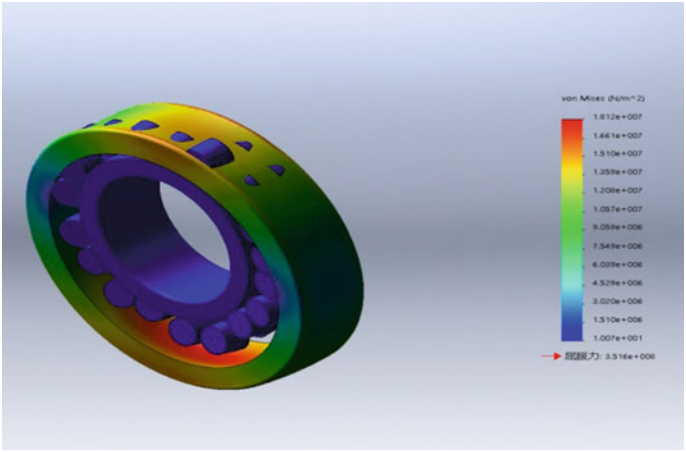


Fig. 6 Stress contour under test load

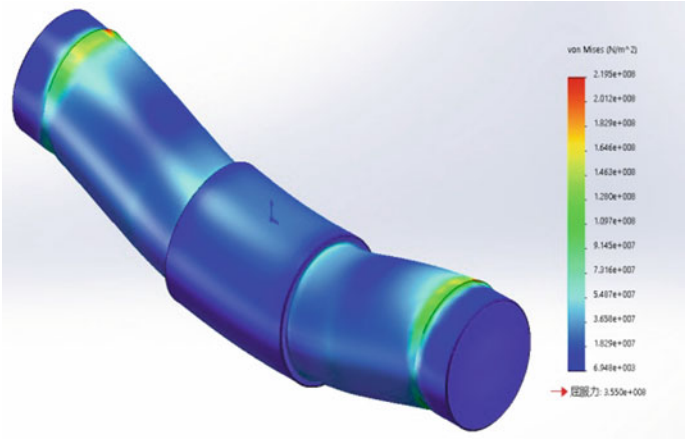


Fig. 7 Stress contour under test load

The material of the sheave shaft is defined as 45 steel, the elastic modulus is 209 GPa, the yield stress is 355 MPa, and the Poisson's ratio is 0.269. Applying load and boundary conditions, the equal and opposite load 18 ton is placed on both sides of the sheave shaft the static stress analysis results is as follows. Figures 7 and 8 are the stress, strain contour under test load. From the graph we can see that the maximum stress is 219 MPa, far less than the yield stress of the material, which can better meet the strength requirements of the large tonnage elevator. Static stress analysis result for maximum resultant displacement 0.048 mm, the maximum strain is 6e−4.

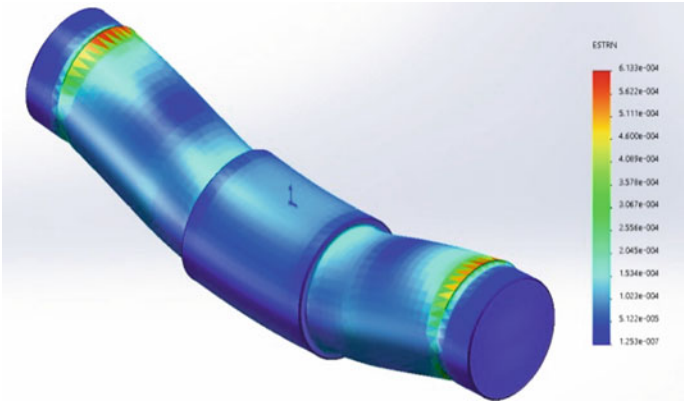


Fig. 8 Strain contour under test load

Through the computer optimization and simulation, the mechanics and the fatigue life of different sheave block are obtained and also the method to improve the service life is obtained in theory. According to the stress analysis and application requirements of large tonnage elevator, study on new materials and structure design of sheave block, reasonably design the connection mechanism and meet the high load requirements using the modern elevator technology.

4 Conclusion

Because large tonnage elevator and vehicle elevator have been more and more widely used in the future, which require higher for traction sheave block, the high efficiency and especially the reliable work should be first considered during the design in the application of large tonnage traction elevator. The sheave structure and relative mechanism need to be optimized to own a better performance through practice and investigation. Good design who has strong vitality and competitiveness will become the ideal substitute in all elevator and will have better market prospect and promotion value. It will also go to the domestic and foreign markets.

Acknowledgments This work is supported by Natural Science Research Major Project of Higher Education Institution of Jiangsu Province (No. 17KJA460001).

References

1. GB 7588-2003 elevator manufacture and installation safety norm
2. GB 25856-2010 freight elevator manufacture and installation safety norm
3. GB/T 31821-15 scrap technical conditions for main components of elevator

4. Jiang X (2015) Research on vibration Control of traction elevator. In: International Industrial Informatics and Computer Engineering Conference (IIICEC 2015), Xian, pp 2144–2147
5. Wang Y (1999) Judgement on effort saving, winding and assembling of sheave block. Harbin
6. Li A (2007) The correct understanding of the effort saving of sheave block. Jiangsu
7. Jian X, Hua Z, Rui Y (2013) Research on intelligent elevator control system. Adv Mater Res 605(11):1802–1805
8. Lianggui P (2006) Mechanical design, 8th edn. Higher Education Press, Beijing
9. Liu H (2010) Mechanics of materials, 5th edn. Higher Education Press, Beijing

A Case Study of Tacit Knowledge Diffusion Object Preference in R&D Teams

Yang Shi

Abstract According to the relevant assumptions of R&D team tacit knowledge target selection, the paper considers "knowledge proximity", "interpersonal proximity", "physical distance proximity", "degree proximity" between members as independent variables, and "tacit knowledge diffusion target selection" as the dependent variable. Through collecting one R&D team tacit knowledge diffusion dates under informal networks environment, it utilizes social network analysis tool and verifies related hypothesis rationality by QAP correlation and regression analysis. The paper finds that "knowledge proximity" and "interpersonal proximity" are the main preferences about tacit knowledge target selection, which will provide some enlightenment on knowledge cooperation and interpersonal interaction in R&D team.

Keywords R&D team · Tacit knowledge diffusion · Select preference Knowledge community

1 Introduction

Generally speaking, tacit knowledge of R&D team comes from two aspects. First, to solve the technical problems of self-learning, through self-study, improve their own professional knowledge and skills, master the R&D process facing new technologies and new methods, in order to keep abreast of knowledge, technological innovation speed. This kind of tacit knowledge can only be achieved successfully through the long time of research and development staff, middle school and middle school. Management guru [1] points out that tacit knowledge like a skill cannot be explained by words. It can only be demonstrated and proved to exist. The main way to learn this skill is to understand and practice. The other way is the teaching of other tacit knowledge workers. A kind of tacit knowledge with

Y. Shi (✉)
Changshu Institute of Technology, Suzhou, China
e-mail: 6910683@qq.com

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_24

inspiration, ideas and thoughts represented only through mutual consultations, in the process of constantly clear these cognitive can produce some new knowledge is not predictable, so as to achieve the transformation and sublimation of tacit knowledge, and this knowledge will be applied to new product development. This kind of tacit knowledge is diffused and transferred mainly through face-to-face communication among members, demonstration, "division with apprentice" and personnel rotation.

From the point of view of social capital, tacit knowledge diffusion is a kind of contact communication in the process of work, life and study. Formal contact working relationship is mainly used for explicit knowledge sharing and transfer based on the related research and development, and personal experience, technical skills, expertise, technical expertise of implicit knowledge only through formal contact between members is not enough, the tacit knowledge sharing and creation of new generation, cannot do without mutual communication and effective interaction between members. Researchers are willing to share tacit knowledge with other members, largely through informal interaction and learning, and subject to the influence of their interpersonal relationships. Informal communication between research and development staff is a resource that can not be imitated, and is also the main form for the whole team to obtain information and knowledge. Especially in recent years and extensive application of the rapid development of computer network, knowledge community presentation rich and colorful in the enterprise (such as intranet, QQ, MSN, BBS, Wiki, etc.) to form a more diversified virtual communication space, some scholars called the virtual community. With the help of more effective, convenient and easy to use knowledge communities, it enriches and expands the social scope and communication space of individuals, and creates the conditions for the diffusion of tacit knowledge.

2 Research Ideas and Questionnaire Design

Using the tools of social relation network analysis, the team knowledge management needs to collect relevant quantitative data, and carries on the empirical test to the social network. The social network requires certain research methods to get the real situation of effective connections among members, which is the basis of data analysis and arrangement.

2.1 Independent Variable Design

"Knowledge proximity" get the data through questionnaires, access to the title file and technical grade archives way. In the questionnaire, we mainly ask questions about module knowledge, structure knowledge and concept knowledge. "Interpersonal proximity" was obtained by questionnaire survey. In the

Table 1 Questionnaire design

Independent variable	Measure item	Investigation method
Knowledge proximity	Do you and he often discuss the principles of knowledge about product development?	Questionnaire answer, data access (title, technical level)
	You and he often discuss the use of different areas of knowledge to solve R&D issues?	
	Do you and he often discuss the main use of R&D products?	
Interpersonal proximity	Do you and he feel similar to each other, the situation is more?	Questionnaire answer
	Do you feel like you can talk to each other honestly?	
	Do you feel that you can get more support and help from each other?	
Physical proximity	Do you feel with each other in response to each other in a timely manner to respond?	Questionnaire answer, on-site observation?
	Do you think it convenient for you to communicate with each other?	
Degree of proximity	Do you think his professional ability is closer to you?	Questionnaire answer, data access (organization chart and job description)
	Do you think he plays the same role with you in the team?	

questionnaire, the main questions about similar background, mutual connection, mutual support and so on. "Physical space proximity" was obtained by means of questionnaires and field observation. In the questionnaire, we mainly ask questions about the convenience and smoothness of the information exchange. "Close to" take the survey and read the organization chart and job description data access data. In the questionnaire, mainly for professional and technical level, social reputation and other related questions (Table 1).

2.2 Dependent Variable Design

Because of the variable "tacit knowledge diffusion, object selection" is mainly obtained by means of questionnaire survey, data consultation and field observation. The main question is, "in your usual interactions, what would you like to share with your members about your research and development experience?". Taking into account the respondents themselves fill out the questionnaire requirements, the survey asked the respondents to fill in not more than 5 members of the code.

3 Example Analysis

3.1 Research Data Sources and Analysis

For the actual situation of knowledge community development team tacit knowledge diffusion, based on a certain enterprise R&D team conducted in-depth research, the specific performance of the team of R&D personnel in the aspects of tacit knowledge exchange and diffusion. Considering the data availability of the tacit knowledge diffusion in the real team, this investigation mainly concerns the knowledge activities of the R&D team members in the QQ group. A total of 70 questionnaires were issued, 52 were recovered, 44 valid questionnaires, the effective rate of recovery was 62.9%.

3.2 The Selection of Tacit Knowledge Diffusion Among Members Based on Knowledge Community

The R&D team members are willing to collect relevant information or tend to share experience and knowledge with other members, will receive the data collation and analysis, generated by social network analysis tools were used to investigate the relationships of enterprise R&D team members between tacit knowledge diffusion and object selection.

Considering some members may involve privacy, the team of R&D personnel number, according to the questionnaire results to whether the exchange of knowledge between members as conditions for the establishment of a 44*44 adjacency matrix, which represents the relationship between member and member. If members are willing to share tacit knowledge with their members, otherwise. The original record into the social network analysis software Ucinet6.1, using the Ucinet software package drawing software Netdraw, rendering knowledge community environment development team tacit knowledge diffusion network map, as shown in Fig. 1.

3.3 The General Situation of Implicit Knowledge Diffusion Preference Among Members

For a comprehensive understanding of the R&D team members' knowledge approach, "knowledge proximity", "interpersonal proximity", "physical proximity", "degree of proximity", answered a questionnaire and on-site observation method, collect the relevant data. Taking into account the relevant issues of the related variables cannot be separated, simple calculation, investigation and other members are assumed to be a member in a "knowledge approach" in the answer very much in line with the assumption that the members, willing to choose some other members

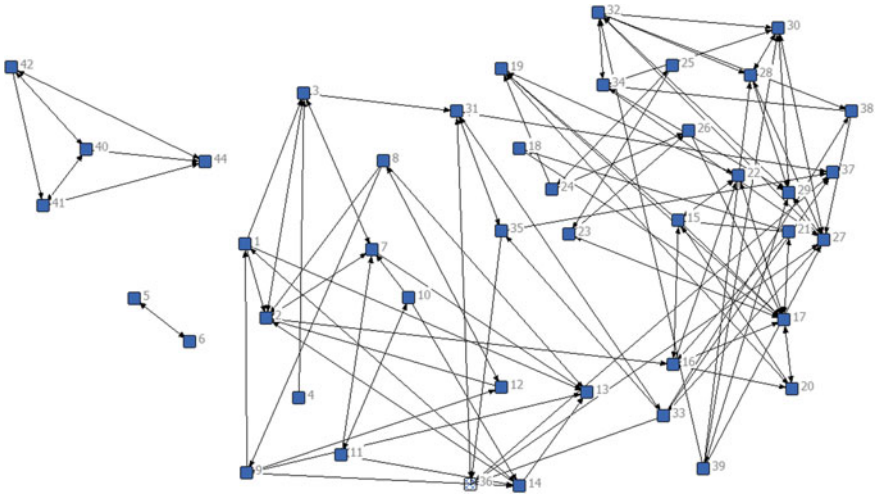


Fig. 1 The implicit knowledge diffusion among members based on informal connections selects the network map

to share knowledge and experience in the relationship matrix denoted as 5, other and so on. According to this train of thought, according to all members of tacit knowledge diffusion objects, select the statistical results of preferences, import social network analysis software Ucinet6.1, draw data matrix Table 3. Considering the limited space, the selection of preference data matrix is omitted based on the implicit knowledge diffusion object of "interpersonal proximity", "physical proximity" and "point degree proximity" (Table 2).

4 Hypothesis Testing of Related Data

According to the theory of social network, the hypothesis testing of the relationship between two kinds of relations is mainly based on the Dyadic QAP method. QAP (Quadratic Assignment Procedure, two times assignment program) is a two square lattice for each value of the similarity comparison method, QAP by re sampling, compare each other lattice array value, the correlation coefficient between the two matrices are given, and the non parametric test of coefficient.

4.1 QAP Correlation Analysis

QAP correlation analysis mainly from the perspective of correlation between the different relationships of the inspection of the assumptions in the statistical sense

Table 2 Selection of preference data matrix of tacit knowledge diffusion object based on knowledge proximity

	1	2	3	4	5	6	7	8	9	10	…	43	44
1	0	0	1	0	0	0	0	0	1	0	…	0	0
2	5	0	0	5	0	0	5	0	3	0	…	0	0
3	0	5	0	1	0	0	4	0	0	3	…	0	0
4	4	3	1	0	0	0	0	5	0	4	…	0	0
5	0	1	0	0	0	5	0	0	0	0	…	0	0
6	0	0	0	0	3	0	3	0	0	0	…	0	0
7	5	4	3	4	0	0	0	0	2	0	…	0	0
8	0	0	0	0	1	0	0	0	5	2		0	0
9	0	0	0	0	1	0	0	4	0	0	…	0	0
10	3	2	0	0	0	3	4	0	0	0	…	0	0
11	1	0	2	3	0	0	0	0	0	5	…	0	0
12	0	0	0	1	0	0	0	3	4	0	…	0	0
13	0	0	5	4	0	1	0	3	0	0	…	0	0
14	0	0	4	0	0	0	5	4	0	0	…	0	0
15	0	0	0	0	0	0	0	0	0	5	…	0	0
…	…	…	…	…	…	…	…	…	…	…	…	…	…
43	0	0	0	0	0	0	0	0	0	0	…	0	3
44	0	0	0	0	0	0	0	0	0	0	…	2	0

correlation, QAP correlation analysis is mainly used “Obs Value”, “Significa”, “Average”, “StdDEV”, Minimum, Maximum, Prop > = 0, Prop < = 0 said. Among them, the “Obs Value” is actually the correlation coefficient between the two matrix; “Significa” refers to a significant level; “Average” is the mean correlation coefficient calculated according to N random replacement; “StdDEV” is the standard deviation; the minimum and maximum number of phase relationship between Minimum and Maximum respectively represent the random calculation in Prop > = 0; the probability of representing these random calculated correlation coefficients greater than or equal to, Prop < = 0 represents the correlation coefficient were calculated is less than or equal to the actual probability of correlation coefficient.

In order to understand the actual situation of tacit knowledge diffusion in R&D teams, the hypothesis of tacit knowledge diffusion was verified. The introduction of R&D personnel between “knowledge approach”, “interpersonal approach”, “physical proximity” and “point of approach” as independent variables, R&D team “tacit knowledge diffusion” as the dependent variable, using QAP correlation analysis to verify the effect of independent variables for variables.

We can be seen from Table 3, the independent “knowledge approach”, “interpersonal approach”, “physical proximity” and “close to nature” and the dependent variable of tacit knowledge diffusion object selection correlation coefficients were 0.579, 0.508, 0.252 and 0.137. According to the calculation results, the “object”

Table 3 Correlation test of tacit knowledge diffusion object selection in knowledge society environment

Independent variable	Dependent variable	Obs Value	Significa	Average	Std Dev	Minimum	Maximum	Prop > = O	Prop <= O
Knowledge proximity	Knowledge diffusion object preference	0.579	0.000	−0.000	0.026	−0.076	0.116	0.000	1.000
Interpersonal proximity		0.508	0.000	−0.000	0.027	−0.077	0.118	0.000	1.000
Physical proximity		0.252	0.000	−0.000	0.026	−0.084	0.107	0.000	1.000
Degree of proximity		0.137	0.000	0.000	0.026	−0.076	0.137	0.000	1.000

and “knowledge approach”, “interpersonal closeness” is the strongest correlation, it shows that the two factors are the main variables affecting tacit knowledge diffusion and object selection in the R&D team.

4.2 QAP Regression Analysis

The purpose of QAP regression analysis is to study the regression between multiple matrices and a matrix, and to evaluate the significance of the determinant coefficient R2. QAP regression analysis considering the independent and dependent variables, mainly the use of “R-square” and “Adj R-Sqr”, “Probability”, “of Obs said #” etc. Among them, “R-square” means “Adj R-Sqr” to determine the coefficient; coefficient of determination is adjusted; “Probability” refers to the one tailed probability; “# of Obs” refers to the number of observation items. According to the relevant theory of social network, the detailed steps of the multiple regression analysis of the conventional long vector matrix and variable matrix variables corresponding to each variable, for each column of the matrix and random permutation, then re calculate the regression, save all the lines and numerical determination coefficient R2. Repeat the above steps to estimate the standard error of the statistic (standard, errors).

It was found that the argument of “knowledge approach”, “interpersonal approach”, “physical proximity” and “close to nature” and the dependent variable of “tacit knowledge diffusion and object selection” to determine the coefficients were 0.335, 0.258, 0.064, 0.019. According to the calculation results, the independent “knowledge approach”, “interpersonal approach” and the dependent variable of “tacit knowledge diffusion and object selection” to determine the highest coefficient, which also proves the R&D team in “close to knowledge” and “interpersonal closeness” is the impact of tacit knowledge diffusion on the choice of the main variables like. The independent variables “physical proximity” and “point degree proximity” are all lower than 0.1 of the dependent variables, which shows that the contribution rate of these two independent variables to the implicit knowledge diffusion object selection may be lower (Table 4).

Table 4 Regression analysis of tacit knowledge diffusion object selection in knowledge society environment

Independent variable	Dependent variable	R-square	Adj R-Sqr	Probability	# of Obs
Knowledge proximity	Knowledge diffusion object preference	0.335	0.335	0.000	1892
Interpersonal proximity		0.258	0.258	0.000	1892
Physical proximity		0.064	0.064	0.000	1892
Degree of proximity		0.019	0.019	0.000	1892

5 Summary

Through the investigation of the R&D team tacit knowledge diffusion, QAP correlation analysis and regression analysis reference the results found in the knowledge community environment variables "knowledge approach", "interpersonal approach", "physical proximity", "close to nature" and the dependent variable of "tacit knowledge diffusion and object selection" correlation. Among them, "knowledge proximity" and "interpersonal proximity" have a significant influence on the choice of tacit knowledge diffusion objects. It shows that whether the tacit knowledge can be effectively diffused and spread within the team is the key to the effective interaction between the members of the real R&D team, whether they can keep each other's professional knowledge and interpersonal relationship.

References

1. Durcker PF (1993) Post capitalist socioty. Butterworth-Heniemann, Oxford, pp 78–89
2. Andrew C, Adva D (1998) Knowledge management processes and international joint ventures. Organ Sci 9(4):454–468
3. Morgan RM, Hunt SD (2005) The commitment trust theory of relationship marketing. J World Bus 40(2):203–221
4. Ancona DG, Caldwell DF (1992) Demography and design: predictors of new product team performance. Organ Sci 3:321–341
5. Baum J, Ingram P (1998) Survival enhancing learning in the Manhattan hotel industry. Manage Sci 44:996–1016
6. Galbraith M, Cohen N (1995) Mentoring: new strategies and challenges: new directions for adult and continuing education. Jossey-Bass, San Francisco

Strategic Framework for Manual Assembly System Design

Logan Reed Vallandingham, Jo Wessel Strandhagen, Erlend Alfnes and Jan Ola Strandhagen

Abstract Manual assembly is necessary for production of many products today, as products are becoming more complex and customer specific. Manual assembly systems are important for competitiveness and are therefore a strategic issue. Previous literature has looked at specific problems, such as line balancing, although a strategic perspective is required to effectively fulfil market requirements. Before designing a manual assembly system, it is important to consider the different impact factors to find the ideal manual assembly system configuration. This study discusses the different impact factors and decisions in strategic manual assembly system design. The purpose of this paper is to develop a framework to assist in strategic design of manual assembly systems. Testing the framework in a case company shows that by designing the manual assembly system according to the identified impact factors, significant reductions of throughput time and non-value adding time in manual assembly can be achieved.

Keywords Manual assembly · Assembly system design

L. R. Vallandingham (✉) · J. W. Strandhagen · E. Alfnes · J. O. Strandhagen
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, Norway
e-mail: logan.r.vallandingham@ntnu.no

J. W. Strandhagen
e-mail: jo.w.strandhagen@ntnu.no

E. Alfnes
e-mail: erlend.alfnes@ntnu.no

J. O. Strandhagen
e-mail: ola.strandhagen@ntnu.no

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_25

1 Introduction

Manual assembly is a critical value-adding activity for many products. Manual assembly performs better than automated assembly for low-volume high-variety production [1]. As products are becoming more complex and customer specific [2, 3], manual assembly will continue to play a dominant role in assembly operations in the future [4].

An assembly system consists of the workers, materials and equipment that produces either components or end products [5, 6]. Assembly system performance has a major impact on competitiveness, and therefore assembly system design is an important strategic decision for companies. A manual assembly system should effectively utilize the workforce to meet customer requirements, and limit disruptions to production [5, 6]. Assembly systems perform better when they are designed according to the product and market characteristics [7]. Assembly lines, for example, are often designed for low cost and high volume production, whereas assembly cells have been used for flexibility. Assembly system design includes many different elements, such as number of workstations, number of workers, and ergonomics considerations [8, 9]. Two key strategic issues, though, are the work organization and the layout type which directly affect the flexibility and volume a system is able to achieve [10].

Many authors have addressed problems within assembly, such as the assembly line-balancing problem [11], assembly line design optimization [12], and sequencing [13]. These assess problems in an existing system, such as an assembly line or cell. There is limited research, however, addressing the strategic aspects related to manual assembly system design. The purpose of this paper is therefore to develop a framework to assist in strategic design of manual assembly systems.

The structure of the paper is as follows: a brief overview of assembly system configurations from literature are presented, where after the impact factors in manual assembly systems are listed and discussed. The factors and previous types are then combined into a framework that will assist in the strategic design of manual assembly systems. A Mid-Norwegian industrial company is presented as a case. The case company produces large, complex electronic equipment, almost entirely with manual assembly. The case company initiated an improvement project to reduce the cost of production, where a new manual assembly system was tested with positive results. We conclude the paper by describing the applicability of the framework and further research opportunities.

2 Assembly Systems

Work organization and layout type have changed over the last century. Assembly lines have been in focus since Ford's development in the early 1900s [14]. At AB Volvo, however, there was a trend to move away from assembly lines in favor of other types of assembly systems [7]. In Japan, assembly cells have become

Table 1 Examples of assembly system configurations

Configuration	Description	References
Line	A product flow where a product is connected by transport technology on a line	[9, 15]
Serial flow	A product flow where every product goes through the same sequence of stations	[7]
Semi-parallel product flow	Product flows that have a common start station, which then branch to include serial and parallel workstations. Products end at different stations	[7]
Parallel product flow	Product flow where some stations that are serial, and other stations are parallel. All products start at the same station	[7]
Different assembly cells	A grouping of stations arranged in a form as to facilitate the work of one or several workers	[7, 15, 16]
Fixed position	Products remain in place, with workers coming to the product to perform the assembly work	[5, 7]

widespread [15, 16]. A list of different examples are shown in Table 1. Examples like these show that certain assembly systems can be more appropriate in certain situations.

Henry Ford's assembly line, and later adaptation by Toyota, are well known and have been studied extensively [14]. Ford's assembly line was designed to reduce the costs of production. There was little room for customization in the assembly line, where workers performed the same operations many times per day.

Semi-parallel and parallel product flow were implemented by Volvo in different plants [7]. These systems were designed to be able to rely more heavily on worker engagement and teamwork. The market changes requiring higher degrees of customization was the main trigger to attempt the implementation of such systems.

Different types of assembly cells have been developed, and have gained interest especially in Japan [15, 16]. Assembly cells can be considered to encompass a few stations, where a worker or workers move around within the cell. These have been implemented to increase the flexibility of assembly systems.

Fixed position (craftsmanship) is considered the original way to perform assembly [14]. This type of assembly is still heavily used in shipbuilding and in assembly of large electronic components [5]. This type of assembly system is used both for practical reasons, as moving large products such as ships can be difficult, but also due to the complexity of the work to be done does not allow for standard work stations.

3 Assembly System Impact Factors and Decisions

In order to gain an understanding of why and how certain manual assembly systems perform better than others do, it is necessary to understand the different impact factors on manual assembly systems. The following main factors are considered:

- Variation in work content [17]
- Handover complexity [17]
- Volume/variety of production [18].

Variation in work content

Some types of assembly systems require a higher degree of worker specialization [19]. A workstation can have a small or large amount of work in the process step ranging from adding a single component, such as a screw, to completing an entire assembly [20]. As the total amount of work increases at a station, so does the variance in operation time at that station. The skill level of workers also has an effect on the resulting complexity in the system, because each worker will complete tasks in a different amount of time [21]. These aspects all contribute to the variation in task completion time.

Handover complexity

In assembly systems, workers have to make choices about which components, tools, and procedures to use in the assembly operations [17]. This complexity is inherent in the variation at their own station, but also of the activities that occurred prior to the station. The variation in operations carried out prior directly affect the complexity of choices of the following operation. The authors suggest this type of complexity to be called "handover complexity".

Volume and variety

The product-process matrix uses volume and variety for determining process structure [18]. Similarly, in assembly the volume and variety affect the type of assembly system that is appropriate [21]. As the volume increases and the variety decreases, automation is more appropriate. However, manual assembly is more appropriate when the production volume is low and the number of variants and customer specifications is high. The authors suggest that within the design of manual assembly systems, volume and variety also affect the most appropriate manual assembly system.

The impact factors discussed above especially affect the following two decisions related to manual assembly system design:

- Layout type [10, 18]
- Number of work groups involved in assembly of one product [7, 10, 20].

Layout type

Layout types can range from fixed position to a single line depending on the volume and variety of production [18]. A higher volume and variety will favor a line-based production; where as a lower volume and variety will favor a fixed position. Between these two extremes, there are several configurations such as parallel flow, which can be ideal in certain situations.

Number of work groups involved in assembly of one product

If the work content at each station in an assembly system varies, having workers follow a single product throughout the entire assembly system can yield better resource utilization [19, 20]. In this way, one worker or work group "chases" the product through the assembly processes, performing all the assembly tasks, without the involvement of other workers. This gives the workers the ability to work at a different pace than other workers in the system, unlike traditional assembly lines. A major issue for "walking workers" is the need for a cross-trained workforce, though, which needs to be considered.

4 Manual Assembly System Framework

A framework has been proposed (see Fig. 1) to show the ideal assembly system based on the related product, process and market characteristics.

The layout type is directly related to the volume and variety of production, as mentioned earlier. Therefore, as a company moves towards a higher volume of production with lower variety, the layout will tend to favor a simplified line flow rather than fixed position. This is in part due to economies of scale that have to be achieved in order to produce in higher volumes [18].

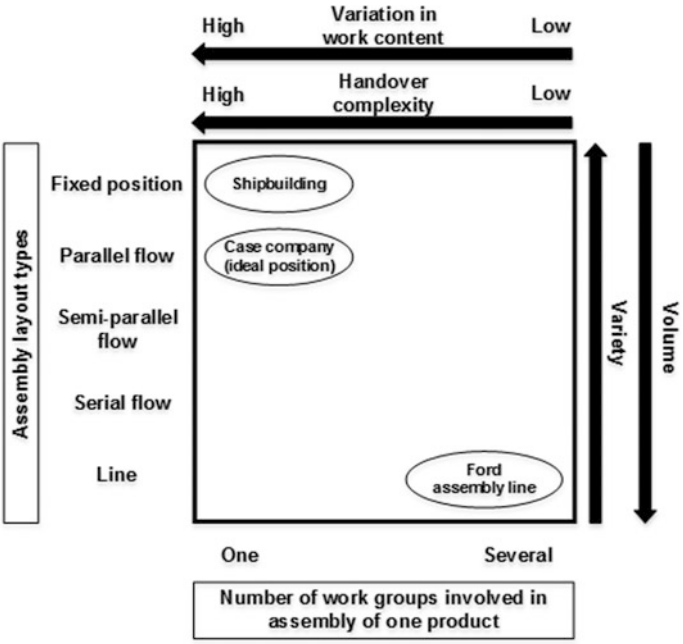


Fig. 1 Proposed manual assembly system framework

Within manual assembly, the question remains whether a worker should operate a single station or be involved in several or all of the workstations. The type of worker dedication is directly related to the handover complexity and the variation in work content. As variation in lead time of each of the process steps increases, it becomes more beneficial to have a single worker follow the product through each of the process steps. A single worker following the product will also be favored if there is an increase in the complexity in handing a product over to the next process step.

Shipbuilding is placed in the top left corner, whereas the Ford assembly line is placed in the lower right. The placement of the case company's product family is shown in the ideal position, and is described in detail in the following section.

5 Case Study—Mid-Norwegian Industrial Company

The case company produces high-tech electronic equipment for large industries, such as maritime and the oil industry. There are five main product families. Due to high costs and low resource utilization the case company initiated an improvement project of one of the main product families.

5.1 AS-IS Situation for the Assembly System

The product family has between 10 and 20 main variants, although they have some specific requirements from the customer. The BOM for the product family consists of approximately one to two thousand electrical and mechanical components.

Generally, the assembly process is manual and consists of workers using simple tools to fit the components together based on the engineering design. The facility had an open floor area for the assembly of the main products, with some tables used for sub-assemblies as needed. Certain areas in the facility are completely dedicated to machines or to making some sub-assemblies and components.

The layout was originally fixed position with a group of workers that assembled the different components, and did the final assembly as they saw fit. The assembly system was not optimal, as there was potential to increase performance.

5.2 Redesign of the Assembly System and TO-BE Situation

After a quantitative and qualitative analysis of the product family, the characteristics were found to have a higher volume and lower variety than the assembly system that they were using indicated. This suggested that the layout should be changed from fixed position to a more flow based configuration. The variation in

work content and the handover complexity however was found to be high, which suggested that the workers should be dedicated to a product rather than to a workstation.

The redesign encompassed creating several workstations for the different process steps of the assembly. It was decided that a small group of workers should move the product through each of these workstations, maintaining responsibility for the assembly of the entire product. Each assembly step had similar components to be produced, but the complexity in handing over the product favored having a dedicated group of workers. Therefore, through the redesign process, the assembly system was designed based on the ideal position as illustrated by the framework (see Fig. 1).

A new layout was partially implemented and a product was assembled (see Fig. 2). A time study was carried out to log the results. The initial results confirmed the improper assembly system. The throughput time for the product was reduced by 38%, due to both reduced material-handling activities and to a simpler material flow through the system. This increased the workers time actually assembling products. Due to the partial implementation, the case company expects up to 50% throughput time reduction in the near future (see Table 2).

Fig. 2 Assembly system pilot with unfinished product shown

Table 2 Expected results of improvement project at case company

	Before	After
Total throughput time (h)	80	40
Value added time (%)	50	80
Non-value added time (%)	50	20

6 Discussion

Manual assembly system design encompasses many different aspects such as specific number of workstations and workers, but it is clear that understanding the strategic implications of a manual assembly system can help lead to improved performance.

Analyzing the strategic choice of which type of system is critical in improving performance and remaining competitive in today's market. The framework, as proposed in Fig. 1, combines different impact factors on manual assembly systems and different decisions that must be taken during the design of a manual assembly system.

The assembly system in use at the case company was a result of years of working in a similar manner, and was ingrained in the workers and management. Upon analysis, it was clear that the fit between the market requirements and the assembly system was not optimal which favored moving towards a more flow based manual assembly system.

7 Conclusion

The assembly of large, complex products often requires a large degree of manual work. The assembly system design has a direct impact on the performance and the ability to utilize the workforce in an effective way. The proposed framework is a step towards theoretical foundation for manual assembly system design, seen from a strategic standpoint. The framework can applied to any company with a high degree of manual assembly.

Further research should aim at testing and refining the framework, as many other elements can affect assembly system design, as well. A quantitative analysis on the impact on performance of different systems for the same products could provide useful insight on how and why performance differs for the different assembly systems.

References

1. Alkan B et al (2016) A model for complexity assessment in manual assembly operations through predetermined motion time systems. Procedia CIRP 44:429–434
2. ElMaraghy H et al (2013) Product variety management. CIRP Ann Manuf Technol 62 (2):629–652

3. Hu SJ (2013) Evolving paradigms of manufacturing: from mass production to mass customization and personalization. Procedia CIRP 7:3–8
4. Butala P, et al (2002) Assessment of assembly processes in European industry. In: Proceeding 35th CIRP-International Seminar on Manufacturing Systems, Seoul
5. DeGarmo EP et al (1997) Materials and process in manufacturing. Prentice Hall, Upper Saddle River
6. Bukchin J, Darel E, Rubinovitz J (1997) Team-oriented assembly system design: a new approach. Int J Prod Econ 51(1–2):47–57
7. Jonsson D, Medbo L, Engström T (2004) Some considerations relating to the reintroduction of assembly lines in the Swedish automotive industry. Int J Oper Prod Manage 24(8):754–772
8. Al-Zuheri A, Luong L, Xing K A framework supporting the design of walking worker assembly line towards improving productivity and ergonomics performance. Scientific Research Publishing (incorporated)
9. Chase RB (1975) Strategic considerations in assembly-line selection. Calif Manag Rev 18 (1):17–23
10. MacCarthy BL, Fernandes FC (2000) A multi-dimensional classification of production systems for the design and selection of production planning and control systems. Prod Planning Control 11(5):481–496
11. Boysen N, Fliedner M, Scholl A (2008) Assembly line balancing: which model to use when? Int J Prod Econ 111(2):509–528
12. Rekiek B et al (2002) State of art of optimization methods for assembly line design. Annu Rev Control 26(2):163–174
13. Bard JF, Dar-Elj E, Shtub A (1992) An analytic framework for sequencing mixed model assembly lines. Int J Prod Res 30(1):35–48
14. Krafcik JF (1988) Triumph of the lean production system. MIT Sloan Manage Rev 30(1):41
15. Miyake DI, de Lima Sanctis R, Banci FS (2010) Shifting from conveyor lines to work cellbased systems: the case of a consumer electric products manufacturer in Brazil. Braz J Oper Prod Manage 4(1):89–108
16. Miyake DI (2006) The shift from belt conveyor line to work-cell based assembly systems to cope with increasing demand variation in Japanese industries. Int J Automot Technol Manage 6(4):419–439
17. Hu S et al (2008) Product variety and manufacturing complexity in assembly systems and supply chains. CIRP Ann Manuf Technol 57(1):45–48
18. Hayes RH, Wheelwright SC (1979) Link manufacturing process and product life cycles. Harvard Bus Rev 57(1):133–140
19. Wang Q, Owen GW, Mileham AR (2007) Determining numbers of workstations and operators for a linear walking-worker assembly line. Int J Comput Integr Manuf 20(1):1–10
20. Wang Q, Owen GW, Mileham AR (2005) Comparison between fixed-and walking-worker assembly lines. Proc Inst Mech Eng Part B: J Eng Manuf 219(11):845–848
21. Al-Zuheri A (2013) Structural and operational complexity of manual assembly systems. J Comput Sci 9(12):1822

The Research and Development of Preventing the Accidental Movement of the Elevator Car Safety Protection Device

ShuGuang Niu, Huiling Liu, Guokang Chen and Haizhou Zhang

Abstract A protective system for preventing accidental movement of an elevator car, comprising a traction mechanism for traction of a car along an elevator guide, the intelligent controller issues a corresponding command to the governor gripper according to the displacement signal, The wire rope and the linkage lever drive the safety gear to fasten the car on the hoistway rail. Able to stop the car when leaving the station but the door is open and the elevator speeding, a high degree of intelligence, simple and reasonable structure. To prevent the traction wheel due to traction wheel failure or traction wheel wear caused by the rope slip brake can not brake, to protect the safety of life and property.

Keywords Car · Accidental movement · Security protection

1 Introduction

With the rapid development of the economy and the deepening of the urbanization process, the number of high-rise buildings in China has increased year by year, and the elevators have become the necessary tools for manned loads in high-rise buildings [1]. The basic principle of the car elevator is the traction drive, the car and the counterweight are connected by the traction wire rope, the hoisting wire rope is hung on the traction wheel, the traction wheel is dragged by the traction motor, so that the trailing wheel Traction rope by the role of friction and movement, and then drive the car lift [2]. In order to ensure that the elevator can be accurate leveling, reliable stop, the elevator also has a brake system, stop the elevator through the

S. Niu (✉) · G. Chen · H. Zhang
Changshu Institute of Technology, Changshu, China
e-mail: niushuguang2000@163.com

S. Niu · G. Chen · H. Zhang
Jiangsu Key Construction Laboratory of Elevator Intelligent Safety, Suzhou, China

H. Liu
Rizhao Maritime College, Rizhao, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_26

brake shoe and the friction between the wheels to stop the car reliable. After prolonged use of the elevator, due to mechanical wear and other reasons, traction wire rope and traction wheel will be different degrees of wear and tear, resulting in reduced drag force, and brake shoes and wear between the wheel will lead to decreased braking force. In addition, the daily use of the overload operation will also lead to lack of traction and lack of braking force. Once the traction and braking force problems, it is possible to cause the car to open the door after the accidental movement away from the station situation, leading to the elevator car and hoist elevator door frame cut the accident. Car accidental movement and elevator over-speed protection is one of the main contents of the current elevator industry research to prevent accidents caused by accidental movement of the elevator, to further improve the safety factor of passengers to prevent personal injury [3].

2 Profile of Car Accidental Movement Protective Device

The first amendment issued by national standard "elevator manufacturing and installation safety norms" GB 7588-2003 On July 2015. Section 9.11 of the revision is described below for the accidental movement of the car: in the case where the door is not locked and the car door is not closed, due to the safe operation of the car, the drive main unit or the drive control system, Failure of the component causes accidental movement of the car from the landing, and the elevator should have means to prevent the movement or to stop the movement. Suspension rope, chain and traction wheel, roller, sprocket, except for the failure of the traction wheel failure, including the sudden loss of traction capacity.

The elevator levels relevels and pre-operates in opening door situation does not need to detect accidental movement of the car when it does not meet the situation 14.2.1.2 and its brake part is the machine match 9.11.3 and 9.11.4.

The standard requirements for accidental movement of the elevator are quite stringent and there must be a device that detects the accidental operation of the elevator at the station, that is, the leveling, the elevator should have the means to prevent the movement or to stop the movement.

3 Design on Safety Protection Device to Prevent the Accidental Movement of the Elevator Car

3.1 Working Principle

For the current requirements of the elevator, designed a safety protection device to prevent the elevator car accidental movement. It uses switching gate electrical signal start anti-protection device when elevator car's landing doors and car doors

are closed, independent of the elevator intelligent controller power supply Into the monitoring state, the detection of car movement state, this time, if the car occurs when the accidental movement, the displacement sensor to the intelligent controller feedback a downward displacement signal, issued a corresponding instruction to the trigger action, trigger wire rope release, So as to promote the safety of the guide rails by means of the safety spring action, and the car is quickly braked on the hoistway guide to stop its descent. If the car moves upwards in a state where the car door is opened, the displacement sensor feeds back an upward displacement signal to the intelligent controller, and the intelligent controller issues a corresponding command to the hoisting rope gripper to prevent the car from continuing rise. When the passenger completely into or out of the car, the car door is closed, the switch door electrical trip switch reset, intelligent controller with power, stop monitoring (Fig. 1).

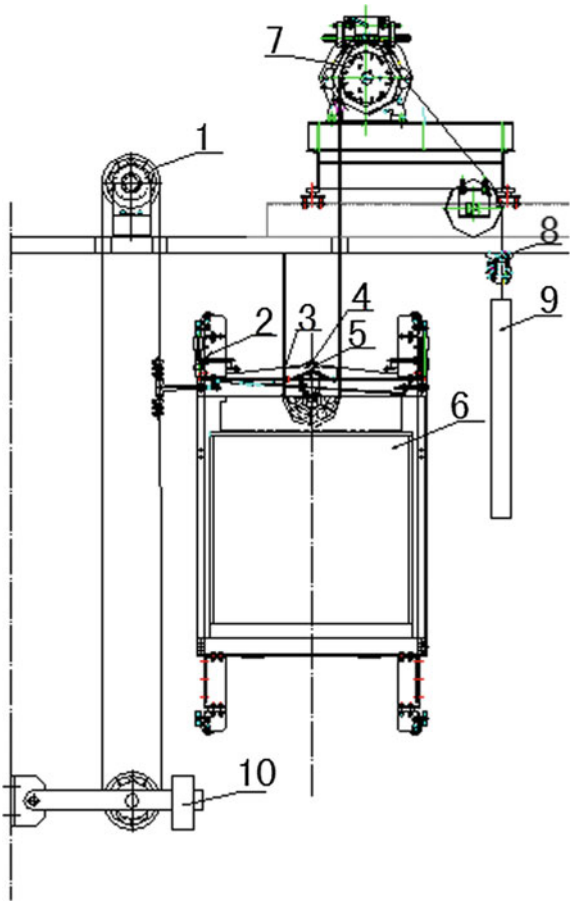


Fig. 1 Schematic diagram of an accidental movement of a safety device for an elevator car. 1, speed limiter, 2 safety gear, 3 trigger wire rope, 4 trigger, 5 monitoring intelligent controller connection, 6 elevator car, 7 traction machine, 8 folder rope, 9 configuration, 10 tension wheel

3.2 Safety Gear

Safety gear and trigger device is the safety protection device of the elevator, when the elevator car accidentally moved and elevator speed, the intelligent controller and speed limiter work, safety gear through the automatic trigger device in the spring under the action of automatic braking, so as to reach the purpose to avoid falling wounding (Fig. 2).

The contact stress calculation of the roller.

Safety gear in the role of the trigger, the release of trigger wire rope and rod, under the action of the spring, the roller up, under the action of the slope, the car parked on the rail.

In the car under the action of gravity, as shown in Fig. 3, the force diagram, the roller relative to the track under the action of the spring is relatively static, inclined block relative to the roller downward movement trend, so the tilt pressure:

$$F_{\text{tilt}} = \frac{1}{n} W / \sin\gamma = \frac{1}{2} \times 7389 / \sin 13^\circ \approx 16424\,N \tag{1}$$

Where

n Safety clamp ramp number
γ The wedge angle of the slant block
W The sum of the car's own weight and load.

Set the wheel by the spring force ignored, the pressure to bear:

$$\sigma_{roller} = 0.418\sqrt{\frac{F_{tilt}E}{lR}} = \sqrt{\frac{16,424\times 210,000}{19 \times 12.5}} \approx 1592\,MP_{\text{a}} \tag{2}$$

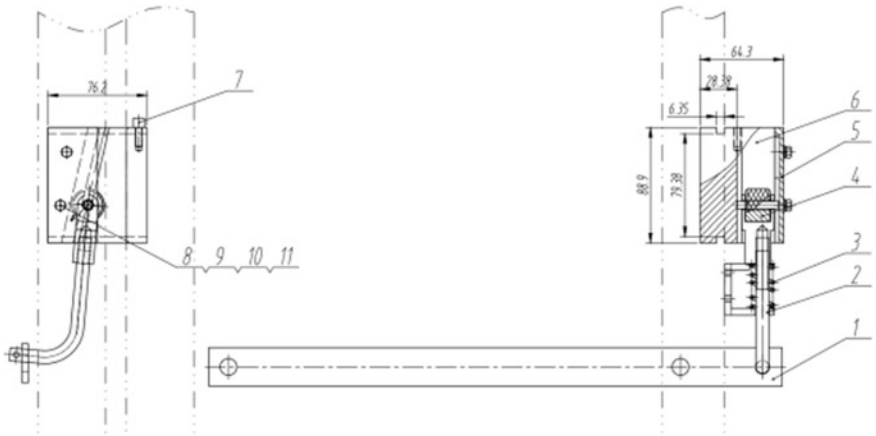


Fig. 2 Schematic diagram of the safety gear

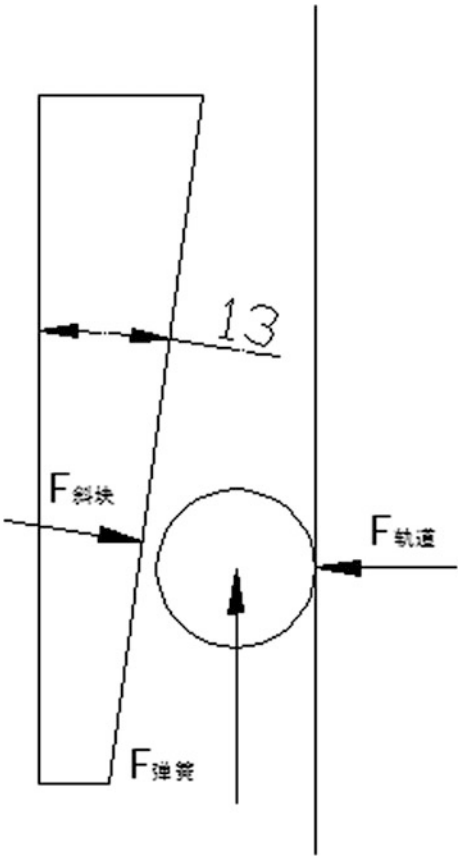


Fig. 3 Force clamp roller force diagram

According to the "mechanical design manual" available, only 40CrV steel, quenched and tempered to meet the requirements, if the elevator weight increases, you can change the wheel into a ramp, you can significantly improve the braking effect to meet the requirements.

Safety gear according to the use of different rails set the appropriate ramp, can be achieved in the event of an unexpected move when the elevator car stop work, to meet international standards. With the manual function of the safety gear, safety gear in any case are effectively clamped on the elevator rails, up to run when no effect; if you need to run down, the operator must manually release the safety gear handle, the safety gear from the premise of the release, The installation platform to run down.

4 Conclusion

Accidental movement of the elevator has been a number of tragedies, causing the public and the industry attaches great importance to many countries in the international community has also enforced the elevator car accidental movement protection requirements. The safety protection device designed to prevent accidental movement of the elevator car and the speed of the elevator can be used to stop the station when the car is opened and leave the station when the car moves out of the station and keep the stop, and the accident is reduced, the degree of intelligence is high and the degree of intelligence is high, The structure is simple and reasonable. To ensure that people enter the elevator car when the personal safety.

References

1. Cai J (2015) Safety protection device designed to prevent accidental movement of elevator car. Equip Manuf Technol
2. Guangwei L (2015) Car accidental movement protection device research. Mech Eng Technol 6:40
3. Guoan L (2012) Elevator open the door under the abnormal movement of the car analysis and prevention. China Spe Equip Saf 11:28

Ultrasonic Sensing System Design and Accurate Target Identification for Targeted Spraying

Xinxue Zhao, Yang Li, Xuemei Liu, Ziru Niu and Jin Yuan

Abstract In order to improve the pesticide application efficiency and reduce the agricultural chemicals lost in fruit trees spraying, an automatic target spraying system of canopy scanning identification based on the sprayer is designed in view of excess spraying and ineffective spray at trees gaps during traditional spray operation. The ultrasonic sensors are installed on the boom of the sprayer under profile modeling spray to detect the crop canopy. Based on the sensor signals, solenoid valves by side are used to control the switch of nozzles at the corresponding position. The experiment of target spray system is worked at the six year-old landscape tree. The results show that the target spray system can effectively save the amount of application pesticides compared with continuous spray. The target spray system can save pesticides 47.8% at the travel speed of 1 m/s for the gap rate 46.7% of landscape tree. Thus, the target spray system of the sprayer has good application value to the pest control of sparse fruit trees and other crops.

Keywords Sprayer · Target spray system · Ultrasonic senor measurement

1 Introduction

In crop precision spraying technology, the crop location and plant size in spray target area are determined whether to implement differentiated pesticide spraying, which can reduce the waste of pesticide, improve the effective utilization rate of pesticide and alleviate the soil pollution. The combination of crop identification technology and pesticide spraying technology, changing from continuous spray to

X. Zhao · X. Liu (✉)
College of Mechanical and Electronic Engineering, Shandong Agricultural University, Tai'an, China
e-mail: lxmywj@126.com

Y. Li · Z. Niu · J. Yuan
Shandong Provincial Key Laboratory of Horticultural Machinery and Equipment, Tai'an, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_27

interval spray technology, which can carry out accurate target spraying become one of the key technologies in precision pesticide application recently [1, 2].

Domestic and overseas scholars have carried out the automatic target application of pesticides in orchard [3], forest [4], protected cultivation [5] and field weed [6]. The research and application of automatic target spraying technology is popular in foreign countries. The detection technology includes image, ultrasonic, infrared and laser which mainly used to prevent and control pests and diseases in orchards such as apple orchard, pear orchard, orange garden, citrus orchard, etc. Compared with continuous spray, the automatic target spray can save 20 to 50% of liquid [7]. In the country, Jin et al. [8] designed automatic target electrostatic sprayer for the orchard and developed a target spray control system. Zhao et al. [5] used the tree image characteristics based on machine vision technology to realize the accurate target application of pesticides. Deng et al. [9] adopted the technique of infrared identification on the side of the plant to achieve the target detection of trees. Lu et al. [10] designed an accurate embedded weeding control system for accurate target spraying which based on ARM7. Zhang et al. [11] used the ultrasonic ranging module for the canopy scanning of fruit trees to obtain the canopy shape. Liu et al. [12] showed that a smart-targeting spraying system on boom sprayer was designed, aiming at satisfying the requirements of spraying for large row spacing and plant spacing.

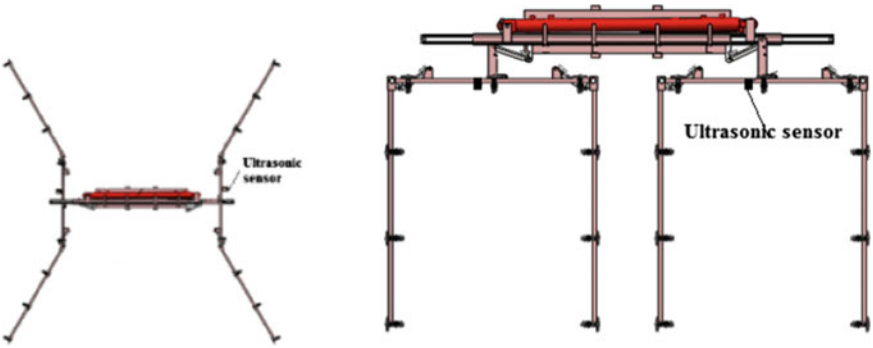
In this paper, a sprayer is used as a platform to design a set of canopy scanning identification target spray system which integrated target ultrasonic detection and nozzle group solenoid valve adjustment. The experiment is carried out on the target spray and continuous spray, and the effect of spraying and spraying on the target spray system is analyzed. By the experiment analyze the effect of spraying and pollution reduction of target spray system.

2 Target Spray System of Sprayer

2.1 Canopy Scanning Identification of Target Spray System

In order to realize the target spray, the ultrasonic distance measurement modules were installed on the boom of the sprayer in profile modeling spray and tunnel spray mode as shown in Fig. 1. In this way can get the distance between the boom and the canopy, and according to the distance information identifying the position of the crop.

At the same time, the nozzles on the sprayer's boom are divided into 2 groups, and the corresponding solenoid valves are installed to control respectively. During operation, according to the canopy shape, adjust the installation angle on both sides of the sensor which corresponds with the canopy. Spray conditions of each nozzle are controlled by the high speed solenoid valve switch and embedded system in real-time. When approaching the target canopy the sprayer automatically spraying, and keep away from the target canopy to stop spraying. Eventually the sprayer achieves the goal of accurate target application.



(a) Profile modeling spray (b) Tunnel spray

Fig. 1 Schematic diagram of spray mode and the sensor's location of target spray **a** Profile modeling spray **b** Tunnel spray

2.2 Hardware Design of Target Spray System

Target detection sensor module used the ultrasonic distance measurement sensor, which compared with laser and infrared is influenced lightly by the intensity of illumination, dust, temperature and humidity of the environment and other factors. The sensor model is MB7052 in XL-MaxSonar-WR, the accuracy is 1 cm, the maximum detectable range is 765 cm, the response distance of the sprayer in profile modeling spray mode is set to 10–200 cm and in tunnel spray mode is set to 10–50 cm. According to the characteristics of spraying operation, the ultrasonic distance measurement sensor is placed on the spray boom to collect the distance between the boom and the crop canopy.

The core of the target spray controller is STM32F103VET6 single-chip computer. The solenoid valve model is 2 W-15 K, the operation time is 100 ms. The data collected by the sensor input to the target system of the controller. Then the target system controller combined with sprayer walking speed to realize real-time target detection. When the target is identified by the recognition program, the solenoid valves which is driven by input and output interface are open, so the corresponding nozzles are sprayed. On the contrary, when the recognition program judges that it is not target, the solenoid valves are closed, so the corresponding nozzle stops spraying. The structures of the target spray system as shown in Fig. 2.

2.3 Target Identification Method

After starting the target system, the sprayer begins walking and collects the distance data in real time. The data is stored in a first-in-first-out queue (FIFO), which the length is N. When the target information is detected, the solenoid valves are opened

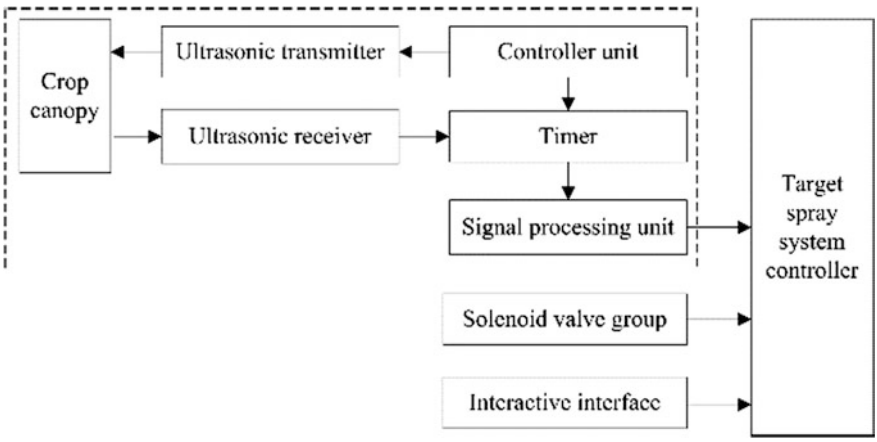


Fig. 2 Structure of the target spray system

and the nozzles begin to spray. End the spraying application, turn off the solenoid valve and start the next target detection. The process is cyclic. In the FIFO, a fixed length sliding window is selected, and the data in the window is analyzed. The length L_w should be able to reflect variability in the data.

$$f_{round}(\frac{w_{\max}}{v_{\min}}f_v) \le L_w \le N \tag{1}$$

where $f_{round}(\ )$ is integral function, $w_{\max}$ is the maximum diameter of the crop canopy, m, $v_{\min}$ is the minimum operating speed of machines, m/s, and f_v is sampling frequency of speed sensor, Hz.

$$N = f_{round}(1.5\frac{w_{\max}}{v_{\min}}f_v) \tag{2}$$

After the width of sliding window is determined, the sliding window moves in the form of one unit step with the acquisition process of the queue data, and the end window always aligns with the latest data acquisition. When the acquisition data is insufficient for the window length during the startup of the machine, fill the window with the method of copying the column header data. The window width calculation ensures that there is at most one mutation point in each sliding window, avoiding the omission or misjudgment of target.

The mean height of the queue is processed, obtain

$$\overline{x} = H - \frac{1}{N}\sum_{i=1}^{N} S_i \tag{3}$$

where H is the maximum value of the queue data, and S_i is raw data collected.

If the mean height of the sliding window is less than the set crop height threshold, it is judged that not target. If the mean height of the window data is greater than the set crop height threshold, the initial judgment is the target, and the position of the first data transition point in the window corresponding to the queue is recorded as the height mutation point. The standard deviation of the data between the mutation point and the next mutation point is further calculated. If the standard deviation is within the standard deviation threshold of the crop distance change, it is determined as the target.

In this paper, the diameter of the crop canopy is in the range of 1.6–2.0 m, the speed of sprayer is 1.0 m/s, the frequency of sensor is 20 Hz, the length of queue is 60 (f_{round}(1. 5 × 20 × 2.0/1.0)). Selecting 20 data after mutation point will not exceed the diameter range of crop canopy, and can better reflect the standard deviation of the crop canopy data.

2.4 Software Design of Target Spray System

The flow chart of the target spray's control system in the sprayer is shown in Fig. 3. After the system is powered on, the sensor is initialized first, and the operating parameters of the sensor and peripheral device are set. And then determines whether to execute target spray. There are two kinds of spraying system model: one is continuous spraying model, and the other is target spray model. If user chooses not to use the target spray system, the solenoid valve has been on opening and continuous spraying. When the user selects the target spray, the ultrasonic sensor information is collected and processed by the acquisition system. If the treated data is crop canopy, the target spray system execute the command to the solenoid valve of nozzle group, through controlling the solenoid valve whether open or not to achieve the target spray. If the treated data is void, the system needs to re-inspection the information and treats it cycling until the spray operation completed.

3 Determine Advance Spray Distance and Spray Time

When the sprayer running at a certain speed, between the ultrasonic sensors detects the target and the nozzle group starts to spray, there will be some time difference. At the same time the spraying should cover the whole canopy area, and then the sprayer needs ahead of a certain distance to start spray and stop spray delay. Therefore, when installing the ultrasonic sensor, along the direction of movement tilt a certain angle. In response time of the target spray system, because of the different traveling distance of the sprayer under different speeds, the advance spraying distance at the corresponding speed changes within a certain range. And the smaller the speed, the greater the spraying distance in advance. Set the delayed spray distance a certain value to calculate the delay time of the corresponding spray

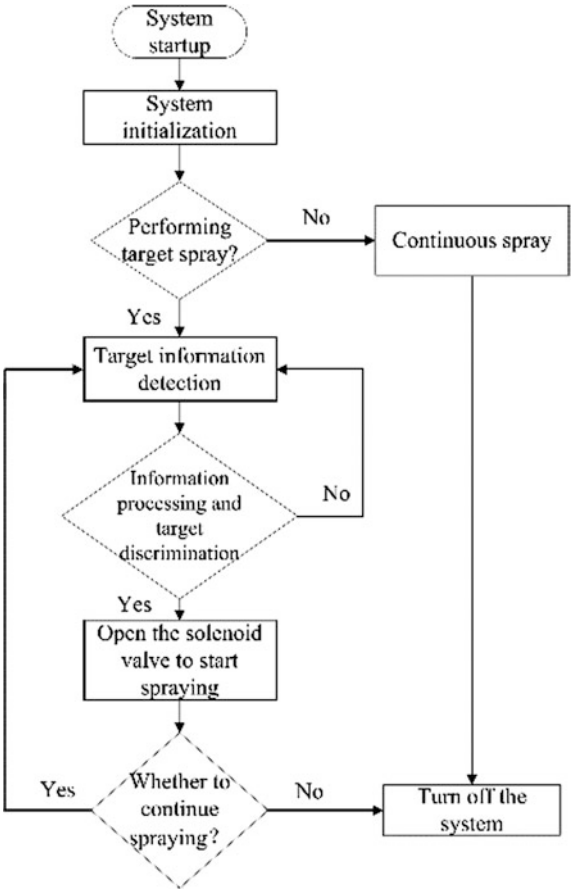


Fig. 3 Flow chart of the control system of the target spray

system. The parameters such as advance spray distance and spray time can be calculated by the next computational formula.

$$l_3 = v_1 t_1, \quad T = \frac{(l_1 + l_2 + l_3)}{v_1} + t_2 \tag{4}$$

Among $t_1 = \frac{l_1}{v_2}$, $\tan\theta = \frac{(l_3 + v_1 t_2)}{l_1}$

where l_1 is the distance between the nozzle and the center of canopy, m, l_2 is the distance of delay spray, m, l_3 is the spray advance distance along the direction of moving, m, l_4 is the crown diameter, m; v_1 is the running speed of the sprayer, m/s, v_2 is the nozzle velocity, m/s, t_1 is the time between the droplet from nozzle to canopy, s, t_2 is the total response time of solenoid valve switch and control system, s, T is continuous spray time, s, and θ is the tilt angle of nozzle along the direction of movement.

Through the advance and delay spray distance setting can ensure that the entire canopy area to obtain effective spray coverage, and adding a certain recognition interval. It can avoid the frequent opening and closing of the solenoid valve caused by the small distance between the trees.

4 Test and Analysis for Target Spray System

The target spray system is tested on the landscape road of Shandong Agricultural University was shown in Fig. 4. The ornamental trees are selected as test objectives and aged for 6 years. The distance between the canopy edges of the selected ornamental tree is obvious, the height of the tree is between 2.5 and 3.0 m, the canopy diameter is between 1.5 and 2.0 m, and the distance between the two trees is 3.0 m, which is suitable for the target spray test. The parameters of landscape trees are shown in Table 1. The driving speed of sprayer is stable at 1.0 m/s. The test medium is homoeothermic water, for the convenience of measurement, the test using unilateral spray. The test has a total of nine nozzles. First, continuous spray, and then the target spray in the test. Before and after each spray measured and record the total amount of water and residual water. The test is repeated 3 times each and the result is shown in Table 2.

It can be seen from Table 2, compared with continuous spray, the target spray can save pesticides 47.8% at the travel speed 1.0 m/s, when the gap rate 46.7% of the landscape tree. It indicates that the target spray can greatly save the amount of application pesticides. The saving rate of pesticides is higher than the gap rate of the landscape trees, because the liquid suddenly releases to the nozzle assembly, when the system detects the target opening solenoid valve on initially. The pressure of the spray unit is low and the spray flow is lower than the normal flow, making the

Fig. 4 Test chart of target spray and signal waveform

Table 1 Parameters of landscape trees

Travel distance/m	Total length of gap/m	Gap ratio/ %	Average spacing/m	Average tree height/m	Average canopy diameter/m
40.5	18.9	46.7%	3.5	2.7	1.8

Table 2 Results of spray test

Spray length/m	Gap ratio/%	Continuous spray		Target spray		Save pesticides/L	Saving rate of pesticides/%
		Spray amount/L	Standard deviation	Spray amount/L	Standard deviation		
21.6	46.7%	6.7	0.08	3.5	0.11	3.2	47.8%

amount of surplus liquid increased slightly. That is, the amount of saving pesticides increased, so the saving rate of pesticides is too large.

5 Conclusion

The target spray was designed based on sprayer, which consisted of XL-MaxSonar-WR series MB7052 ultrasonic sensors, multi-nozzle, corresponding solenoid valve and embedded control system. The system used ultrasonic sensor to collect the canopy target information, and used embedded control system to adjust the switch of the nozzle solenoid valve in real time to realize the target spray of fruit trees and other crops.

Install the distance sensor in the front of the nozzle at 40 cm. At a speed of 0.4 m/s, there is 1 s to ensure that the system's solenoid valve moves. The target spray system in the determination of the target start edge, according to the test set the delay time, and advance the target 2 cm distance began to spray, leave the target 2 cm distance to stop spraying, so that the canopy can be fog shrouded to ensure that the application effect.

Compared with continuous spray, the target spray system could effectively save the amount of application pesticides. The target spray system could save pesticides 47.8% at the travel speed of 1 m/s for the gap rate 46.7% of landscape tree. Thus, the target spray system of the sprayer has good application value to the pest control of sparse fruit trees and other crops.

References

1. Liu XM, Yuan J, Zhang XH et al (2012) Development and experiment on 3MQ-600 type air-assisted boom sprayer with air-deflector. Trans CSAE 28(10):8–12
2. Brown DL, Giles DK, Oliver MN et al (2008) Targeted spray technology to reduce pesticide in runoff from dormant orchards. Crop Protection 27:545–552
3. Jejčič V, Godeša T, Hočevar M et al (2011) Design and testing of an ultrasound system for targeted spraying in orchards. Strojniški vestnik-J Mech Eng 57(7–8):587–598
4. Hočevar M, Širok B, Jejčič V et al (2010) Design and testing of an automated system for targeted spraying in orchards. J Plant Dis Prot 117(2):71–79

5. Cao ZY, Zhang JX, Geng CX et al (2010) Control system of target spraying robot in greenhouse. Trans CSAE 26(Supp2):228–233
6. Yin DF, Chen SR, Mao HP et al (2011) Weed control system for variable target spraying based on fuzzy control. Trans Chinese Soc Agric Mach 42(4):179–183
7. Stajnko D, Berk P, Lešnik M et al (2012) Programmable ultrasonic sensing system for targeted spraying in orchards. Sensors 12(11):15500–15519
8. Jin X, Dong X, Yang XJ et al (2016) Design and Experiment of Target Spraying System of 3WGZ-500 Sprayer. Trans Chinese Soc Agric Mach 47(7):21–27
9. Deng W, He XK, Zhang LD et al (2008) Target infrared detection in target spray. Spectro Spectral Anal 28(10):2285–2289
10. Lu J, Mao HP, Chen SR (2009) Embedded weed control system for target precision spraying based on ARM7. J Agric Mechanization Res 12:76–79
11. Zhang L, Zhao ZX, Yu L et al (2010) Positioning algorithm for ultrasonic scanning of fruit tree canopy and its tests. Trans CSAE 26(9):192–197
12. Liu XM, Li Y, Li M et al (2016) Design and test of smart-targeting spraying system on boom sprayer. Trans Chinese Soc Agric Mach 47(3):1–10

Fault Location in Power System Based on Different Modes of Traveling Wave and Artificial Neural Network

Chenglei Liu, Ke Bi and Rui Liang

Abstract Intelligence algorithm benefits most of diagnosis and repair working. Fault location in distribution power system is vital for the electronic power supply constantly. A novel fault location method based on ANN is used for complex power system. The amplitude of transient initial traveling wave is used to select the measuring points which form the fault scope. The arrival time of zero mode and aerial mode traveling wave are calibrated by wavelet transform. In each level of detail coefficient, the time difference is different, which is relate to the fault distance and used for the input of ANN. Then, the accuracy fault location can be obtained in fault scope using ANN. PSCAD/EMTDC is used to verify the effectiveness of proposed method.

Keywords Fault location · ANN · Wavelet transform · Traveling wave

1 Introduction

With the rapid development of economic, electronic supply play an important role in daily life. But the fault in power system is common. Such as grounding fault, lighting strike, etc. As a result, fault location in distribution power system is vital for the electronic power supply constantly. Now, computational intelligence technology has deeply penetrated in industry. Intelligence algorithm benefits most of diagnosis and repair working [1].

C. Liu (✉) · R. Liang
School of Electrical and Power Engineering, China University of Mining and Technology, Xuzhou, Jiangsu, China
e-mail: 1109420374@qq.com

K. Bi
School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou, Jiangsu, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_28

Traditional fault location method has some unavoidable defects. The popular fault location methods are impendence-based method and traveling wave-based method. Impendent-based fault location does not need high sample frequency and can be realized by existing measurement equipment. However, this method is susceptible to line parameters, fault resisters, load swing, etc. Impendent-based methods are unsuitable for long transmission lines which with many changeable factors [2–4]. Traveling wave based methods are immunity from above factors and have great accuracy and robustness, which need higher sample frequency than impendent-based methods and need traveling wave detectors [5–8]. Those method has a great performance on long transmission line but not good for distribution power system which with many unbalance loading and small current neutral grounding system. Computation intelligence who has more sensitivity and robustness will play great performance on fault location in distribution power system. Literature [9] shows a comprehensive review on method for voltage sag source location and proposes another novel method using a robust support vector machine (SVM) in which many features are extracted, based on previously described methods. Then, the source location by machine learning technique is discussed and the SVM with the linear, polynomial, and radial basis function (RBF) kernels are applied along with optimal genetic search, which will benefit fault location as so. Literature [10] proposes fault section locating for distribution network with DG based on improved ant colony algorithm. A regional processing method is used in the paper, which divides the distribution network into several independent regions, then the fault locating algorithm is used in these independent regions. The regions without DG connection are eliminated if fault current is not detected. So the complexity of computation of fault-section location is reduced. Ant colony algorithm is a good optimization algorithm of swarm intelligence. This intelligence method in power system diagnosis and repair achieve great performance. There no double that computational intelligence is an effective method for fault location in power system.

In this paper, a novel Artificial Neural Network (ANN) method for fault location in distribution system has been proposed. Due to the frequency dependent characteristic, zero-mode traveling wave velocity is different from the aerial-model traveling wave velocity. The time difference of different mode arrive at measuring point is link to the distance that the traveling wave transform. The amplitude of zero-mode traveling wave head is used to ensure the fault area. The time difference of zero mode and aerial mode arrival (TDZA) is trained by ANN to locate the fault location. The wavelet transform is used to calibrate the time stamp and calculate the energy of different measuring points.

The reminder of this paper is organized as follows. Effectively fault scope is made sure in Sect. 2. Section 3 introduction ANN and the fault location algorithm. Section 4 gives the case study and test results. Conclusions of this work are present in Sect. 5.

2 The Process of Fault Area Identification

When a fault happens on the line, a transient traveling wave emerges at fault points and travels to each side of the line, until all the power grid. As shown in Fig. 1.

The Karrenbauer matrix (1) is used to transform the coupled phase A, B, C to independent mode 0, 1, 2 where mode 0 is the ground mode and the mode 1, 2 are the aerial mode in engineering.

$$\begin{bmatrix} U_0 \\ U_1 \\ U_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} * \begin{bmatrix} U_A \\ U_B \\ U_C \end{bmatrix} \tag{1}$$

According to the traveling wave expressions (2), the amplitude of measured signals is attenuating exponential with distance.

$$I(x,t) = \sqrt{2} I^{+} e^{-\alpha(\omega)x} \cos\left(\omega t - \beta(\omega)x + \phi_{+}\right) + \sqrt{2} I^{-} e^{\alpha(\omega)x} \cos(\omega t + \beta(\omega)x + \phi_{-}) \tag{2}$$

So, far away from the fault points, the measured signals' amplitude is lower than that of the closer. Compare with aerial mode traveling wave, the zero mode traveling wave head attenuation is severe and hard to be detected in the distance, which highlight the area of around fault location and can be used to make sure the fault area. Where the amplitude is high, the fault location is nearby this measuring points.

Due to the frequency dependent characters of lines, the attenuation coefficient and phase coefficient are different for each signals. The transient traveling wave heads are composed of different frequency signals. Wavelet transform is an efficient tool to separate different frequency bands and calibrate the arrival of transient traveling wave head. As a result, the amplitude of transient traveling wave head is composition of different frequency bands.

Db6 wavelet is used in the paper to transform the transient signals into 6 levels. The modulus maximums of each detailed coefficients level are a1, a2, a3, a4, a5. We defined the amplitude of transient traveling wave head measured is:

$$A = a_1 + a_2 + a_3 + a_4 + a_5 \tag{3}$$

For n measuring points in the power grid, there are n measuring amplitudes A_1, A_2, …, A_n. The three largest of A_1, A_2 …, A_n are the measuring points make up the fault area, which will used for the fault location as shown in the last paper detailed.

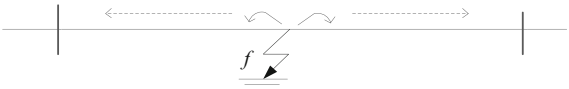


Fig. 1 Transient traveling wave

3 Fault Location by ANN

3.1 ANN

ANN exhibit excellent normalization and generalization capability, robustness, immunity to noise, and system parameter estimation errors. Nowadays, many researchers study the computational intelligence technology and proposes ANN-based fault location methods, which present a better performance in the presence of fault resistance and noise. Moreover, since ANN-based fault locators do not require an exact knowledge of the power system configuration; they have been extensively used in power system applications. Figure 2 shows the architecture of a typical feed-forward multilayered neural network which consists of an input layer, one or more hidden layers, and an output layer. the x_n is number of inputs. $\omega_{kn}^{(1)}$ and $\omega_{mk}^{(2)}$ are the weights applied to the second and third layers, respectively. The number of hidden layers and neurons in the layers is subjected to the problem researched. In this paper, the neurons in the output layer is one, which means the fault distance from the measuring points. The data received from outer are trained by the ANN. The neuron calculates a weighted average of them using the summation function and then uses an activation function to compute the output. Then, a non-linear relationship process between the input and output data can be built $\omega_{m,k}^{(2)}$.

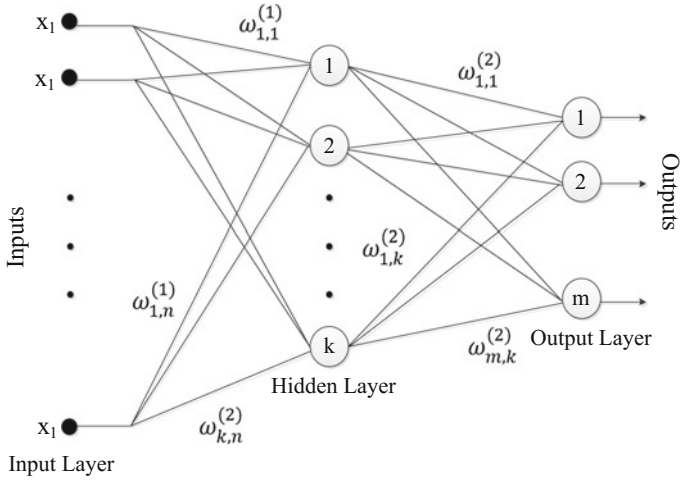


Fig. 2 Feed-forward multilayer ANN

3.2 *Fault Location Algorithm*

In Sect. 2, three measuring points which form the fault area can be obtained. The zero mode and aerial model traveling wave can be measured clearly in this three points. Due to the different traveling path, the velocities of zero model and aerial model is different. The velocity of zero model is always slower than that of aerial model. Wavelet transforms the transient traveling into several levels. In each level, the time difference of zero mode and aerial mode is different, which is link to the fault distance and signals frequency.

Figure 3 gives a time difference of zero mode and aerial mode in different levels of wavelet transform. Table 1 gives the time difference of two modes in different fault distances. From it we can see that the time differences are different in different levels. The time difference is related to the distance from fault point and measuring point.

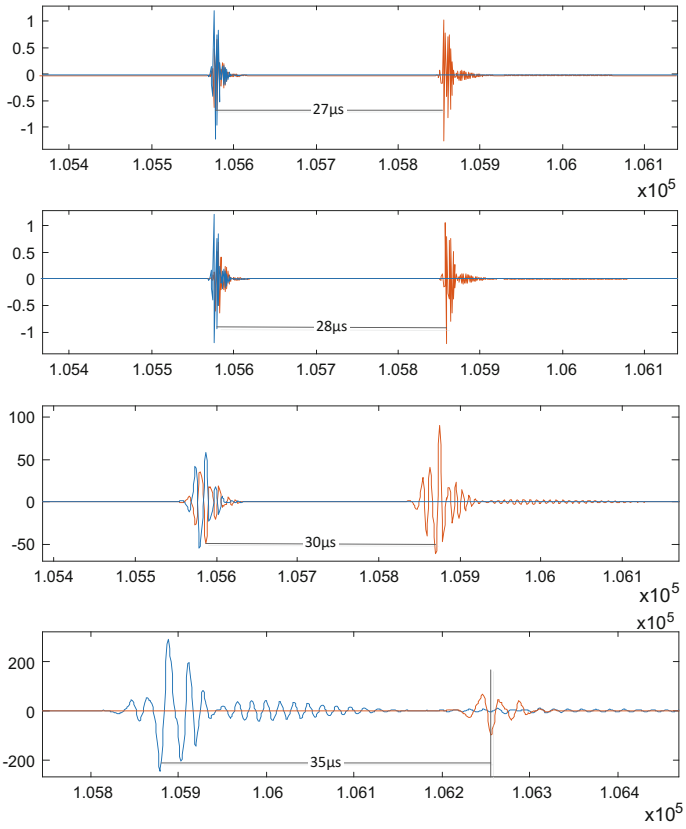


Fig. 3 Time differences in different levels

Table 1 Mode velocity and arriving time at different fault distances in d1 level

Fault distance (km)	Arrival time of zero mod (s)	Velocity of zero mode km/s	Arrival time of aerial mode (s)	Velocity of aerial mode km/s	Time difference (μs)
100	0.25339	2.950×10^5	0.25334	2.994×10^5	5
200	0.25681	2.937×10^5	0.25668	2.994×10^5	13
300	0.26029	2.915×10^5	0.26002	2.994×10^5	27
400	0.26378	2.903×10^5	0.26336	2.994×10^5	42

So, this method takes the time difference of different levels as the input of ANN. The modulus maximum of each levels (d1, d2, d3, d4) as the initial voltage traveling wave arrival time. Finally, the corresponding time differences are $\boldsymbol{s} = [\Delta t_1, \Delta t_2, \Delta t_3, \Delta t_4]$, which is the input of ANN. The output of ANN is the fault distance from the measuring points. The unit of output is km, the unit of input is μs. The process of normalization is necessary:

$$x_k = \frac{x_k - x_{min}}{x_{max} - x_{min}} \tag{4}$$

In other words, there are altogether four neurons in input layer and one neuron in output. Formulation (5) has been used to make sure the number of neurons in middle layer:

$$m = \sqrt{n+p} + \alpha \quad (\alpha = 1, 2, 3, \ldots, 10) \tag{5}$$

n is the number of neurons in input layer and p is the number of neurons in output layer. Adjust α from 1 to 10 to find the best value of m. In this paper, the value of m is 10. The structure of ANN is $4 \times 10 \times 1$ for this fault location algorithm. *Tansig* function and *purelin* function in MATLAB are used for the hidden layer and output layer respectively. Adaptive learning algorithm is adopted, whose learning rate is set to 0.01, maximum training times is 1000 and the convergence error is 10^{-5}. From simulation results, the convergence accuracy is achieved after 20 times training.

For each measuring point, the fault distance can be obtained. Then, severe points on adjust branch can be obtained. Based on Sect. 2, three measuring points are used for fault location. Assume the distance from measuring point A_1 is x_1, A_2 is x_2, A_3 is x_3. As shown in Fig. 4, the intersection of point f is the fault point. Due to the measuring errors, the point f may not exist. In this situation, the intersection point f can be calculated by the average distance from each measuring point distance. For measuring point A_1, the fault point from A_1 is l_{x_a1}. For measuring point A_2, the calculated fault distance far from A_1 is l_{x_a2}. For measuring point A_3, the calculated fault distance far from A_3 is l_{x_a3}. Then, the accurate fault distance from Bus A_1 can be calculated by:

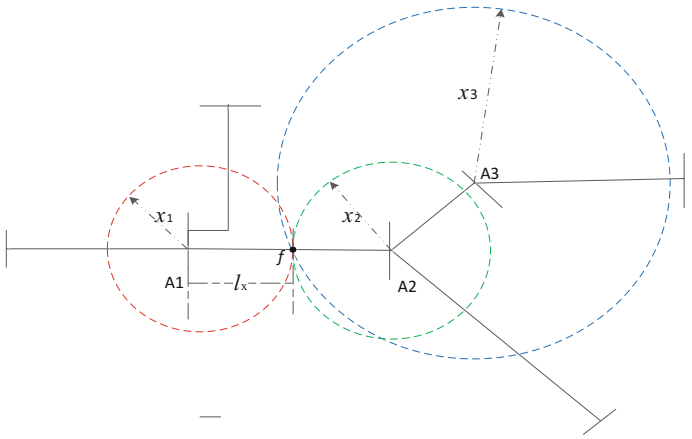


Fig. 4 The fault location result of intersection

$$l_x = (l_{x\_a1} + l_{x\_a2} + l_{x\_a3})/3 \tag{6}$$

4 Performance Evaluation

The proposed fault location methodology is evaluated through PSCAD/EMTDC. The simulation model is based on the standardized IEEE-30 Bus system. Figure 5 is the diagrammatic sketch of IEEE-30 Bus system.

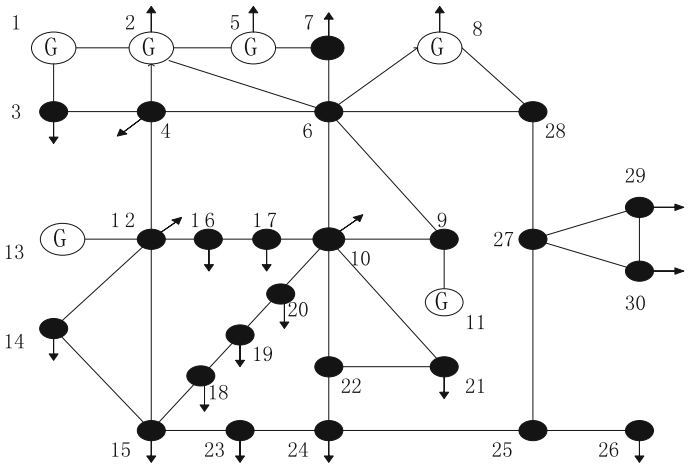


Fig. 5 IEEE-30 Bus system

Different faults with different positions, fault angle, fault resistance has been simulated. The process of calculated is running in MATLAB.

We set a single-phase grounding fault on the line of 10–17, which is 30 km from bus 10. The amplitudes of transient traveling wave head measured at each buses are calculated by (3). The result is shown as Fig. 6. It is clear that the bus 10, 16, and 17 have the largest amplitude, which is selected as the fault scope and used for fault location.

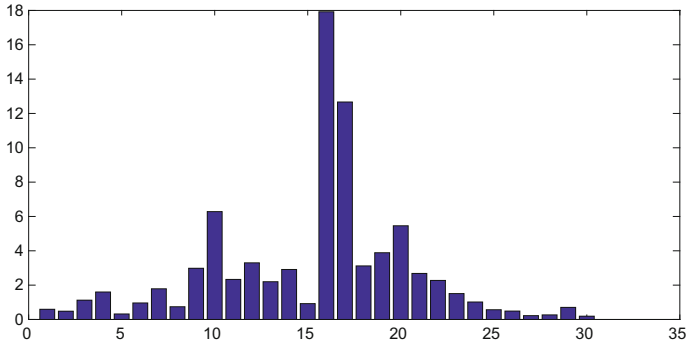


Fig. 6 The transient amplitudes of traveling wave

Table 2 The fault location results

Fault distance/km	Fault angle/(°)	Fault resistance/(Ω)	Fault location/km	
			Double-ended	Proposed
10	30	0.1	10.21	9.88
	60	10	10.21	9.82
	90	100	10.21	9.78
	90	300	10.21	10.37
36	30	0.1	36.34	36.89
	30	10	36.34	35.11
	60	100	36.34	36.92
	90	300	36.34	36.75
55	30	0.1	55.16	55.05
	60	10	55.16	54.58
	60	100	55.16	54.74
	90	300	55.16	55.21
72	30	300	71.80	71.78
	60	10	71.80	72.12
	90	100	71.80	72.21
	90	0.1	71.80	72.05

Then the trained ANN is used for the fault distance from the three measuring points and the (6) is used for the accurate fault location. The fault distance obtained by proposed algorithm is 30.016 km. The absolute error is 0.036 km, which meet the required of practical engineering.

In order to verify the accuracy and adaptability of proposed fault location method, different faults are set on line 10–17, with different position, fault angle and fault resistance. Table 2 gives the result of fault location in different situations. Comparing with traditional doubled-end method, which need synchronization, proposed method even has a better location results!

5 Conclusion

This paper proposes a novel fault location method in complex power system, which do not need the synchronization and the reflected wave identified. The wavelet transform for the modulus maximum is reliably. First, according to the energy of transient current traveling wave, the three points which near by the fault points is chosen. The transient energy is almost linear to the fault distance, which can be the feature characters for the fault area. So, this method is robustness for the complex power system.

Using the computational intelligence—ANN, proposed method has a better accuracy. The time difference of each layer corresponding to different frequency bands. In different frequency bands, the velocity of traveling wave is different, the proposed method considers the situation that frequency dependent characteristic, which will increase the accuracy of fault location. The result of fault location can satisfy the demand of practical engineering.

References

1. Ünal F, Ekici S (2017) A fault location technique for HVDC transmission lines using extreme learning machines. In: 5th International Istanbul Smart Grid and Cities Congress and Fair (ICSG), Istanbul, Turkey, pp 125–129
2. Bains TPS, Sidhu TS, Xu Z, Voloh I, Zadeh MRD (2017) Impedance-based fault location algorithm for ground faults in series capacitor compensated transmission lines. In: IEEE transactions on power delivery, vol 99. pp 1–1
3. Naidu O, George N, Pradhan D (2016) A new fault location method for underground cables in distribution systems. In: First international conference on sustainable green buildings and communities (SGBC), Chennai, pp 1–5
4. Bazargan M, Maleka K (2017) Fault location algorithm for identifying a faulted section in a meshed multi-terminal HVDC grid. In: 13th IET international conference on AC and DC power transmission (ACDC 2017), Manchester, pp 1–6
5. Lopes FV, Neves WLA, Femandes D Jr (2013) A traveling-wave detection method based on park's transformation for fault locators. IEEE Trans Power Delivery 28(3):1626–1633

6. Alireza Ahmadimanesh A, Shahrtash SM (2013) Transient-based fault-location method for multiterminal lines employing s-transform. IEEE Trans Power Delivery 28(3):1373–1380
7. Liang Rui, Wang Fei, Guoqing Fu (2015) Wide-area fault location based on optimal deployment of the traveling wave recorders. Int Trans Electr Energy Syst 26(8):1667–1672
8. Feizifar B, Haghifam MR, Soleymani S (2011) Application of continuous wavelet transform for fault location in combined overhead line and cable distribution networks. Electr Electron Eng 14(2):142–146
9. Mohammadia Y, Moradia MH, Leborgneb RC (2017) A novel method for voltage-sag source location using a robustmachine learning approach. Electr Power Syst Res 45:122–136
10. Tao W, Yang G, Zhang J (2016) Fault section locating for distribution network with DG based on improved ant colony algorithm. In: IEEE 8th international power electronics and motion control conference

Review of Technology for Strengthening Effect of Fe Element in Al Alloys

De-Qin Sun, Zi-Qiang Zhou, Su Zhu and Pei-Jun Wu

Abstract When the Fe content in Al alloys was high, Fe-rich phases usually were precipitated in the forms of thick flaky or needle-like, which can greatly reduce the mechanical properties. One research area was how to reduce Fe content in Al alloys, and another was that changing the morphology of Fe-rich precipitates in the form of thick flaky or needle-like into regular morphologies such as rod-like, massive, even spherical, and its grain size was refined by some technologies such as alloying, cooling control, heat treatment, et al. and this would greatly improve the mechanical properties of Al alloys. The research of Al alloys with high Fe content was already one hotpot in the fields of novel Al materials study and had made some progress.

Keywords Fe-rich phase · Dilution method · Compositional overcooling Rapid cooling

1 Introduction

Al alloys had many superior properties, such as lower density, higher strength-to-weight ratio, excellent electrical and thermal conductivity, good corrosion resistance, and so on. These superior properties made it to be an ideal structural material. For example, a large number of Al alloys were used in the aircraft carrier, such as 1019 t Al alloys were used in USS CVA62 Independence, and 450 t were used in USS CVA65 Enterprise [1].

But most of the Al alloys was relatively soft, its hardness was low, so its ability of wear resistance was limited. This was one of the reasons why Al alloys were not

D.-Q. Sun (✉) · S. Zhu · P.-J. Wu
School of Chemical& Materials Engineering, Changshu Institute of Technology, Changshu 215500, China
e-mail: sundeqin@cslg.edu.cn

D.-Q. Sun · Z.-Q. Zhou
Jiangsu Key Laboratory of Recycling and Reuse Technology for Mechanical and Electronic Products, Changshu 215500, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_29

widely used in engineering field. Taking the automotive manufacturing industry as an example, with the worsening of environmental pollution and the increasingly serious of energy crisis, vehicle lightweight had become one of the themes of social development, and the use of metallic materials of little density, such as Al and Mg alloys, was the most important technology. For the most Vehicle components, relatively higher demanding upon the mechanical and technical capacities were needed, not only high strength and toughness, good corrosion resistance, abrasion resistance, heat resistance and so on, but also good machinability and hot-working performances. As the increasing of the Fe content in Al alloys would improve the wear resistance and high temperature strength, so the Al–Fe series Al alloys with Fe content of more than 1% had attracted the widely attention in Al processing research and technological study.

2 The Role of Fe Element in Al Alloys

Traditionally, Fe was generally considered harmful element in Al alloys, especially in wrought Al alloys, which generally required the Fe content under 0.3%, even less than 0.1% in some high performance alloys [2]. This can be analyzed by the Al–Fe phase diagram (rich Al side) [3]. As shown in Fig. 1, the solid solubility of Fe in Al base was very small, the maximum amount of solid solution is only about 0.04% at the eutectic temperature (928 K), and almost 0 at room temperature. In Al alloys with a certain content of Fe, Fe element was mainly precipitated in the form of metallic compound Al_3Fe, which showed in the shape of flake or needle-like, under the conventional cooling condition; if also contained Si element in Al alloys, Si would be involved in the forming of Al_3Fe precipitations, and produced the brittle flake precipitations such as α-Fe_2SiAl_8 or β-$FeSiAl_3$, which existed in the Al matrix or at the grain boundary. The existence of these precipitations seriously produced the fragmentation effect to Al matrix, and produced Local stress concentration and increased the brittleness, resulted in significantly the decrease of the plasticity and toughness [4]. For example, when w(Fe + Si) >0.14% in 2519 alloy, more coarse brittle compounds AlCuFe was precipitated at the grain boundary with the increase of w(Fe + Si) content, and the strength of the alloy was obviously decreased [5].

But for the casting Al alloys, the demands for the plasticity, toughness and so on were not strict, and the requirements of the Fe content also were not strict. In conventional casting Al Alloys of the Al–Si series, increasing the Fe content would still have an adverse effect on the mechanical performance, significantly decreased the strength, toughness, corrosion resistance and other mechanical properties of casting Al alloys; simultaneously, increasing of Fe content could reduce the fluidity and mold-filling capacity of Al–Si Melts; furthermore, more brittle particles appeared in Al casting, made the machining processing to be more difficult [6]. However, increasing the Fe content was benefit for the mechanical properties in high temperature, the main reason was that the Fe-rich phase was a compound with good heat stability, which existed in Al matrix in the form of square or fine acicular,

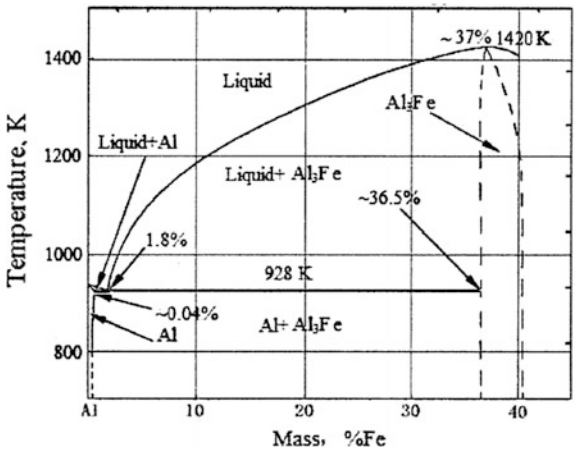


Fig. 1 The Binary Phase Diagram of Al–Fe (Al-rich parts)

and network-like distribution, led to the dispersion strengthening effect on the Al base, improved the tensile strength at the high temperature. At the same time, the hard Fe-rich phase which dispersed in Al matrix, increased with the increase of Fe content, and increased the ability of the wear resistance [7]. For die casting Al alloys, also needed more higher Fe content, mainly based on two reasons: (1) the increase of Fe content was beneficial for taking the casts out from the mould. Generally, in the die casting process, Al alloys was very easy adhered on the surface of the mold. When the Fe content was more than 0.6%, the status of mold sticking was obviously reduced, so the Fe content was generally controlled in the range of 0.6–1%; this would help to increase the ability of avoiding the surface welding during the Al liquid was filled into the mould; (2) For die casting alloys in the Al–Si series, when the Fe content was in the range of 0.6–0.8%, the changes of fracture strength and elongation were not obvious with the increase of the Fe content changed, but when the Fe content was over 0.8%, the fracture strength was greatly reduced from 230 MPa to about 200 MPa. This was due to the higher cooling rate in the die casting process, the grown-up of flake or needle-like Al_3Fe brittle precipitations was restricted, and showed relatively small particles and more uniform distribution, which were benefit for its mechanical properties; while the Fe content is more than 0.8%, the morphology of Fe-rich precipitations had not been modified. Therefore, the Fe content was generally lower than 1% for die-casting Al alloys [8].

3 Technical Research Work of Eliminating Harmful Effects of Fe in Al Alloys

Actually, whether the wrought or cast Al alloys, the Fe content was controlled in a relatively lower content range in the production process. Especially for wrought Al alloys, the Fe content was strictly controlled to a lower level in order to prevent the

production of the precipitation of flake rich Fe phase. Therefore, in order to ensure Al alloys to achieve good toughness and deformation ability, a lot of research works of eliminating the harmful effects caused by the high content of Fe, mainly embodied in two aspects: (1) reduced the content of Fe in Al alloys; (2) improved the precipitates patterns.

3.1 Technology of Reducing the Fe Content in Al Alloys by Means of Feedstock Pre-processing or Melt-Treatment

It was relatively easy to control the Fe content in the primary Al production process, but for scrap recycling process, because of the complexity of Al alloy raw materials and recycling processing, often mixed with some materials containing iron, so the Fe content was often far beyond the expected chemical composition. In order to avoid excessive Fe elements to mixed into, the first work was that excluding the steel materials as much as possible from the Al scraps.

(1) The most widely applied technology was magnetic separation, which used for separating the iron and steel materials from Al scraps. Its principle was: making use of the difference of magnetic properties, the magnetic separation equipments were installed in the raw material transferred line, the iron and steel components were separated from the Al scraps due to the steel materials were adsorbed and conveyed to given area under the action of an electromagnetic field. Due to the equipments were relatively convenient, electromagnetic process was more extensive applied in domestic and abroad [9]. (2) Melting technology. For those components, which mechanically combined by Al and steel parts, it was necessary to disassemble and sort them by hand, because It was difficult to separate them by electromagnetic process. Utilized the difference of molten point, placed the scrap materials included steel inserts into a special melting furnace, and heated over the Al molten point, the steel inserts were picked out of the Al melt [10]. (3) Gravity separation technologies, which utilized the density difference of different kinds of raw materials scraps, or by means of existed in layer due to the different settling velocity in the liquid medium, or was thrown to the different areas by the force of the high-speed conveyor belt [11]. (4) Eddy current separation technologies, which utilizing the difference of the conductivity and density between different metals, placed scrap materials within an alternating magnetic field or electrostatic separation equipment, and the different scraps were thrown to different areas due to the different electromagnetic induction force, and achieved the separation of different scraps [12]. (5) In recent years, some new techniques, such as color classification technology, chemical composition analyzer, etc. was developed. The principle of color classification was that, carried out chemical corrosion to the Al scraps, and distinguished different materials through color or graphics analysis of the scarp particles by computer; The technology of chemical composition analyzer included

X ray fluorescence analysis (XRF), light emission spectroscopy (OES) and laser induced breakdown spectroscopy (LIBS), etc., which were brought into the material sorting field, and made more accurately separation effect [13].

Nowadays, the sorting capacity of above Al scraps separation technologies was limited, there were still a large number of Fe elements were brought into the smelting process. Therefore, the research on the removal of excess Fe in Al scrap melt was also one of the research hotspots. (1) The most commonly used was the dilution method [14], its principle was that, added a certain amount of pure Al ingots into the scrap melt, and reduced the Fe content of the Al scrap melts. This needed a lot of pure Al ingots to be consumed, and greatly improved the process cost. (2) Elemental neutralization technology, a process by adding Mn, Cr and other elements in the Al melt with higher Fe content, Mn and Cr could promote to form the primary Fe-rich precipitations of high melting point, such as α-Al_{15} (Fe, Mn)$_3Si_2$, α-Al_{15}(Fe, Mn, Cr)$_3Si_2$, etc. Then removed the primary Fe-rich precipitations out of Al melting by means of gravitational, filtration, electromagnetic and centrifugal separation technologies, and reduced the Fe content [15]. (3) Added the fluxes composed of boride, etc. into the Al melts with a certain Fe content, the chemical reaction occurred between the boride and Fe element, and created the compounds Fe_2B which had a high melting point [16], easier to segregate together in Al melt and formed larger particles. These compounds were brought into the molten slag and removed from Al melt by the melt purification process. (5) The special properties of RE elements had been attracted the wide attentions, especially in the studies of functional materials. In recent years, the research works for the RE elements in removal Fe element by forming intermediate compounds involved Fe and the effect of modification and purification to the Al melt, had become one of research hotspots.

3.2 *Technology of Morphology Controlling of Fe-Rich Precipitated Phases*

In die casting Al alloys, the Fe content was about 0.6–0.8%. But under the technological conditions of higher cooling rate, the growth of Fe-rich precipitates was restrained, showed the morphology of more smaller particles and more uniform distribution. This was helpful to improve the strength and toughness of Al alloys. Therefore, modified the morphology of Fe-rich precipitates into the shape of fine rod, bulk and spherical, the harmful effects would transform into a beneficial effect of the dispersion strengthening or dislocation pinning to the Al matrix, and improved the mechanical performance of Al alloys. In fact, there was a very good example in cast iron materials, which the precipitated morphology of graphite showed in form of flake graphite, the tensile strength of gray cast iron was generally lower than 250 MPa, however the precipitated morphology of graphite showed in form of spherical graphite, the tensile strength of ductile iron was generally over

600 MPa, and the elongation ratio increased to 5–10%. This was proved that modification of the precipitates morphology would greatly improve the mechanical properties, such as the tensile strength and the toughness, etc.

The research work about the modification of Fe-rich precipitates morphology in Al alloys mainly in two technologies: neutralized by addition of alloying elements and control of cooling rate in solidification process.

The study on the modification of Fe-rich precipitates morphology by adding stabilizing elements, included: (1) Created non-spontaneous nucleation particles by adding alloy elements and changed its growth mode of the primary $A1_3Fe$, achieved the optimization in the morphology and distribution status of primary Fe-rich precipitates. For example, adding Zr element in Al alloys could generate $A1_3Zr$ phase, which may be accomplished as the crystal nuclei in the precipitated process of the primary $A1_3Fe$, obviously reduced the needle-like precipitates, and the morphology of Fe-rich phases showed mostly regular particles. (2) Improved the constitutional overcooling in front of solidification shell, created more the crystal nuclei of primary $A1_3Fe$ in the Al alloy melting. For example, added a small amount of Ca element into the Al–Fe alloys, played the roll of improving the constitutional overcooling in front of solidification shell, the Al_3Fe particles became smaller and regular. (3) Added Mg, Co and RE elements, such as Y, Sc et al. could prevent the diffusion of Fe element in Al melts, and the grown-up of the Fe-rich precipitates could not carry on. For example, Mg and Y, Sc and other RE elements could be gathered on the surface of primary Fe-rich precipitates, hindered the diffusion of Fe element, and led to the increasing of crystallization overcooling, prompted the increase of crystal nuclei by grown up in branches, improved to refine the primary nucleation. The existence of Sc element could also make the interfacial energy balanced in different directions of the Fe-rich nucleation basement, and made the Fe-rich grown up well-distributed in different directions, formed a more regular precipitated morphology. The addition of Co elements brought up different functions, Co could be almost completely dissolved into the Al_3Fe phase, became the diffusion barrier of Fe atoms, thus inhibited the particle coarsening of Al_3Fe phase. (4) By means of elemental neutralization technologies, modified the composition structure of Fe-rich precipitated phase, and improved the morphology and uniform distribution of Fe-rich precipitates. The addition of Mn element into Al–Fe alloys led to form the (Fe, Mn) Al_6 precipitations in shape of short rod-like and massive, and reduced the deleterious effect of Fe-rich precipitates.

Many research achievements also had been obtained in technological study for the modification of the Fe-rich precipitation morphology, such as control of cooling rate, electromagnetic stirring, etc. (1) Combined the process of extrusion casting and Mn elemental neutralization technologies, the Fe content in Al–Cu alloys could be increased to 1%. According to the Clasius-Clapeyron law, the increase of stress in the course of solidification led to the raising of eutectic temperature, improved the overcooling, thus increased nucleation rate of the crystalline grains. When improved the extrusion pressure during solidification process, the Fe-rich precipitation morphology changed from needlelike shape to short rod-like, and the Fe-rich phase was refined; This also increased the binding force of the interface between

Fe-rich phases and Al matrix, and reduced the cracking tendency. Under the process conditions which the cast pressure was 75 MPa, the elongation of Al alloys was about 13.5%; There was the same results when utilized the technology of centrifugal casting, which could make the primary $A1_3Fe$ phase more refiner of the outer layer. (2) The precipitated phases in Al–Fe alloys were effectively refined by the semisolid extrusion process, its principle was that: after semi-solid extrusion deformation, existed unstable high free energy in Al alloys, easy made the matrix and the Fe-rich precipitates to occur recovery and recrystallization, and refined the Fe-rich precipitates. The results showed that, after semi-solid extrusion process, the size of the primary Al_3Fe phase was about 25% of that in ordinary cast process. (3) Treated by electromagnetic stirring technology to the Al alloy melt with high Fe content, could locally change the microstructure of liquid metal because of the Joule heating effect, improved the solidification overcooling and increased the rate of the nucleation rate; meanwhile, broken up the grown Fe-rich precipitates and refined the Fe-rich precipitates under the role of the stirring effect. (4) Carried on the rapid cooling treatment to Al–Fe alloys, the lamellar spacing decreased sharply of Al–Fe primary eutectic phases, the morphology of Al_3Fe precipitates changed from coarse lamellar into fine regular shape and dispersed distribution. (5) According to the principle of manufacturing the particle reinforced composites, added the Fe powder into the Al melt with the help of stirring treatment by plasma spray process, and poured into moulds within 5 min; obtained good morphology of Fe-rich phase, the Fe-rich precipitates displayed more fine particles.

4 Development of Al Alloys with High Fe Content

After modified the morphology of Fe-rich precipitates in Al alloys through different technologies, the mechanical properties of Al alloys with Fe high content (over 1%) were significantly improved. For example, the tensile strength, elongation and corrosion resistance would be increased more than 50%. Therefore, the research and development of Al alloys with high Fe content had been caused enough concerns in Al manufacturing and technology fields and had made some achievements.

Al–Fe–V–Si (FVS0812) alloy were a kind of novel superalloy developed in 1986 by Allied-Signal Co. of the US, manufactured in process of combination of rapid solidification and powder metallurgy technology (RS/PM). In this alloy, created a certain amount of dispersed α-$Al_{12}(Fe,V)_3Si$ particles which had excellent thermal stability, the size was 40–50 nm in shape of spherical, strengthened the matrix and grain boundaries, thus this alloy processed good properties included not only low density, high strength, but also good thermal stability and high fracture toughness, etc., could replace Ti alloys and be used as novel heat resistant structural material in service conditions of high temperature below 400.

In China, the study of Al–Fe–V–Si Al alloys was developed by means of injection formation technology. The manufacturing process was that: the chemical composition was Al-8.5Fe-1.3 V-1.7Si(mass, %), melted in Medium Frequency

induction furnace, sprayed into cylindrical billet of diameter 125 mm, and then hot extruded at 410. The alloy possessed the compact structure, fine grain and uniform microstructure. At room temperature, the tensile strength was achieved about 415 MPa, the yield strength of 345 MPa or so, and the elongation was more than 20%. Meanwhile, it also possessed good mechanical properties at high temperature.

In addition, the study on Al–Fe alloys by means of the chemical composition adjustment, control cooling, deformation processing and other technologies, was also made some progress. The bulk nano-crystals Al–Fe alloys with uniform microstructure and fine grains, were achieved by the process of mechanical alloying and spark plasma sintering (MA-SPS). At room temperature, the compressive strength reached up to 1000 MPa, and the one-time plastic strain deformation ratio was more than 30%; the round billet of Al Si Fe Cu Mg La Ce alloy with 10% Fe content was obtained by vacuum spray deposition, which could be manufactured into many kinds of components, such as pistons, connecting rods, cylinders, valves, gears, etc., and could be widely used in automobiles, machinery, metallurgy, electrical industries; the manufacturing process of mine cables by Al–Fe–Mg–Zr alloys with 0.2–1.1% Fe content was that: the Al–Fe–Mg–Zr bar were made by the process of melting, casting, rolling, and then the cables were achieved by drawing, stranded, sheath treatment, etc. The cables possessed good properties, such as the resistivity of no more than $0.028264\Omega\bullet mm^2/m$, the elongation of over 10%, and had excellent performances of high strength, high conductivity and good bending ability.

5 Conclusion

In the process of conventional Al manufacturing technologies, the Fe-rich precipitates were easily formed in shape of coarsening flake or acicular, which was seriously harmful to the mechanical properties, when the Fe content in the Al alloys was higher. But the increase of Fe content could make Al alloys have good high temperature performances and wear resistance. Therefore, Al alloys with high Fe content possessed its unique advantages in some special industrial applications, could meet more stringent conditions such as engineering, aerospace, shipbuilding, automobile parts, building templates and other engineering fields. At present, the study on Al–Fe alloys focused on the manufacturing process, such as solidification control, heat treatment deformation, etc., and modified the shape of Fe-rich precipitates by increasing the crystal Nuclei, refined its grains and promoting its dispersed distribution;

Nowadays, the process of Al scrap recycling was becoming increasingly important. Using the characteristics of contained more Fe element in Al scrap materials, developed some Al–Fe alloys and its manufacturing technologies, such as taking reasonable alloying, melt treatment and solidification control, etc. This could effectively reduce the manufacturing cost, and was favorable to the development of Al–Fe series alloys.

References

1. Wen-jie W (2013) The application status and perspective of alloys for high performance and advanced naval vessels. Mater Rev 7:98–105
2. Xiao Y-q (2004) Practical Handbook of aluminum processing technology.Metallurgical Industry Press, Beijing
3. Zhen-ping Zhou (2008) Research on Melt Treatment and solidification characteristics of AI–Fe Alloy. Shenyang University of Technology, Shenyang
4. Ceschini Lorella, Boromei Iuri, Morri Alessandro et al (2011) Effect of Fe content and microstructural features on the tensile and fatigue properties of the Al–Si_{10}–Cu_2 alloy. Mater Des 36:522–528
5. Li H-Z, Liang X-P, Zhang X-M et al (2007) Effect of Y content on microstructure and mechanical properties of 2519 aluminum alloy. Trans Nonferrous Metals Soc China 39 (3):527–531
6. Taylor John A (2012) Iron-containing intermetallic phases in Al–Si based casting alloys. Procedia Mater Sci 1:19–33
7. Seifeddine Salem, Johansson Sten, Svensson Ingvar L (2008) The influence of cooling rate and manganese content on the β-Al_5FeSi phase formation and mechanical properties of Al–Si-based alloys. Mater Sci Eng A 490(1):385–390
8. You-dong Xu, Cheng Chen, Jun-jun Zhang (2015) Influence of iron contents on the mechanical properties of Al–Si die-cast alloy. Guangdong Chem Industry 42(1):28–29
9. Chen J, Zhao J, Zhang M (2011) Metal recycling technology. Chemical Industry Press, Beijing pp 35–50
10. Majidi O, Shabestari SG, Aboutalebi MR (2007) Study of fluxing temperature in molten aluminum refining process. J Mater Process Technol 182(1):450–455
11. Fan C, Long S-Y, Li C et al. (2012) The status and development of separating and sorting technology on Aluminium scrap. Light Metals (07):61–64
12. Gao A-J, Wang G, Qu X-L et al. (2015) The research on impurities separating and sorting technology in the pretreatment process of the aluminum scrap recycling. Recyclable Res Circular Economy 8(2):33–36
13. Cho Thet Thet, Meng Yi, Sugiyama Sumio et al (2015) Separation technology of tramp elements in aluminium alloy scrap by semisolid processing. Int J Precis Eng Manuf 16 (1):177–183
14. Zhang Lifeng, Gao Jianwei, Damoah Lucas Nana Wiredu et al (2012) Removal of iron from aluminum: a review. Miner Process Extr Metall Rev 33(2):99–157
15. Zhang Lifeng, Wang Shengqian, Dong Anping et al (2014) Application of electromagnetic (EM) separation technology to metal refining processes: a review. Metall Mater Trans B 45 (6):2153–2185
16. Wang Y-B, Yi G-B, Yang X-J (2014) Effect of $Na_2B_4O_7$ on iron removal influence for recycled aluminum Alloy.Foundry 63(10):1024–1027

Characterization the Fatigue Life of the Flame Thermal Spray Parts Applying the Power Function Energy Method

Zhi-Ping Zhao, Chao Wang and Xin-Yong Li

Abstract In order to predict fatigue life of components which are made by 40 Cr steel substrate and thermal spray coated with Ni60A film. From experimental investigation on combined bending and torsion low cycle fatigue performance of samples, the effect of remelting time were compared. Hence, reasonable remelting time which satisfy the low cycle fatigue experiments was obtained. Related material parameters which used in life prediction model were fitted by elastic-plastic stress-strain relationship and least square method which were supposed by Remberg-Osgood. Then, the fitted material parameters were evaluated by the results from low cycle fatigue experiments. Hence, the equation of power function energy method which could apply on flame thermal spray components was established. Meanwhile, the combined bending and torsion fatigue life of flame thermal spray components could predicted under various low load conditions.

Keywords Thermal spraying component · Combined bending and torsion fatigue test · Fitting parameters · Life predict equation · Fatigue life prediction

1 Introduction

Remanufacturing components with thermal spraying process are going to be served in serious working condition. Hence, study on fatigue life prediction of this kind components is significant to construct fundamental theory of remanufacturing and benefit to develop life prediction model. In the study of fatigue life, a successful prediction is an extremely complex task [1]. According to various structure, material, production, loading and environment, the fatigue life prediction of structure would be analyzed differently. Thus, fatigue life prediction was concerned

Z.-P. Zhao (✉) · C. Wang · X.-Y. Li
School of Mechanical Engineering, Changshu Institute of Technology, Changshu Jiangsu Province, China
e-mail: 781244066@qq.com

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_30

and studied by many engineering and academic researches. A serious of corresponding fatigue life prediction models were developed by researchers [2–6].

Currently, study on fatigue life could be divided to high cycle fatigue (more than 105 cycles to failure) and low cycle fatigue (less than 105 cycles to failure). In high cycle fatigue situations, test on material performance is commonly characterized by applying controlled stress loading. And in low cycle fatigue situations, test on material performance is commonly characterized by applying controlled strain or energy loading. Hence, if the characteristic of high cycle fatigue could be expressed by energy method also, there would be no boundary existed in the analyze between high cycle fatigue and low cycle fatigue. With analyzing experimental results from 3 different materials and according to the fatigue parameter α which was presented by Mc Dowell and Miller [7], a uniform three parameters power function energy method(3USE) which could predict both high cycle fatigue and low cycle fatigue was developed by Xu et al. [8].

In this study, the 3USE method was employed to predict low cycle fatigue life of thermal spraying components. By constructing fatigue life prediction model of the components, a practical meaning on estimating fatigue life of thermal spraying components was offered by this study. Based on previous researchers [9–11], least square method was applied to fit the material parameters employed to elastic-plastic stress-strain relationship and power function energy method. Moreover, a mathematical model to evaluate the parameters in power function energy method was developed. Finally, with applying the calculated material parameters from low cycle fatigue performance experiments, the equation of power function energy method which could apply on flame thermal spraying components was established.

2 Fatigue Life Prediction Model with Applying 3USE Method

The uniform three parameters power function energy method (3USE) which could employ both high cycle fatigue and low cycle fatigue is expressed as below:

$$N_f = (\Delta a - \Delta a_0)^m = C \tag{1}$$

$$\Delta_a = \Delta a_e + \Delta a_{in} = \frac{\Delta\sigma\Delta\varepsilon_e}{2} + f\left(\frac{1}{n'}\right) * \frac{\Delta\sigma\Delta\varepsilon_{in}}{2\pi} = \frac{(\Delta\sigma)^2}{2E} + f\left(\frac{1}{n'}\right) * \frac{\Delta\sigma}{\pi}\left(\frac{\Delta\sigma}{2K}\right)^{1/n'} \tag{2}$$

$$f\left(\frac{1}{n'}\right) = 3.85\sqrt{\frac{1}{n'}}\left(1 - n'\right) + \pi n'$$

$$\frac{\Delta\sigma}{2} = K(\frac{\Delta\varepsilon_{in}}{2})^{n'}$$

$$\Delta\sigma = \sigma_{max} - \sigma_{min}$$

where, Δ_a is fatigue parameter amplitude, Δ_{a_0} is fatigue limitation, m and C are constants. In the situation of simple fatigue nonlinear strain region, $\Delta\varepsilon_{in}$ could be replaced as Δ_{ε_p}, E is elastic modulus, K is hardening constant and n' is hardening exponent.

3 Combined Bending and Torsion Fatigue Life Test

3.1 Test Material

The shape and size of specimen were decided by reference [12] with 40 Cr steel substrate and Ni60A coating film. The chemical components (wt%) of self-fluxed powder alloy Ni60A are 13. 7% Cr, 2. 96% B, 4.40% Si, 2. 67% Fe, 0. 60% C, and the remaining part is Ni.

Before thermal spraying, the specimen was carried with turning process. Then, the coating surface was polished by SiC sandpaper and aluminum oxide (Φ0.3 μm). After polishing, alcohol and acetone were used to remove grease. Finally, oxy-acetylene flame was employed thermal spraying process. The first stage: pre-treatment of surface → pre-heating → pre-dusting → dusting. The second stage: remelting → cooling. The remelting process was carried out immediately at one side of dusting surface since the temperature of specimen was raised greatly to 200–400 °C. In order to study the effect of remelting time, 4 kinds of different remelting times were investigated. Case 1–4 have the duration of 2, 5, 10 and 12 min separately. After the remelting process, specimens were cooling down to room temperature in asbestos powder. The coating thickness of each specimen was around 1 mm.

3.2 Test Method

The combined bending and torsion fatigue life was tested by SHIMADZU) EHF-EM 100 kN electric-hydraulic test machine. The effect from corrosion was neglected since the experiments were carried out at neutral environment. A regular sinusoidal stress was applied with 1 Hz frequency and R = −1 loading ratio. The maximum loading applied on specimen was ± 800 N and the minimum loading applied on specimen was ± 400 N.

4 Results and Discussions

4.1 Fatigue Life of Specimen

According to the relationship of fatigue loading and fatigue life in Fig. 1, a significant effect from remelting time was observed. The 2 min case has a shortest fatigue life. Then the fatigue life is observed to be increased in 5, 12 and 10 min case. Which means the 10 min case has the longest fatigue life.

4.2 Data from Fatigue Life Test

From the investigation of fatigue life on remelting specimen, a longest fatigue life was showed out by 10 min case. Thus, parameters related to the 3USE method were fitted in the case of 10 min. According to experimental study, the low cycle fatigue performance date of 10 min case is showed in Table 1.

4.3 Least Square Method to Fit Parameters in Remberg-Osgood Equation

From the relationship in Eq. 2, Δa could be calculated by elastic modulus E, hardening constant K and hardening exponent n'.

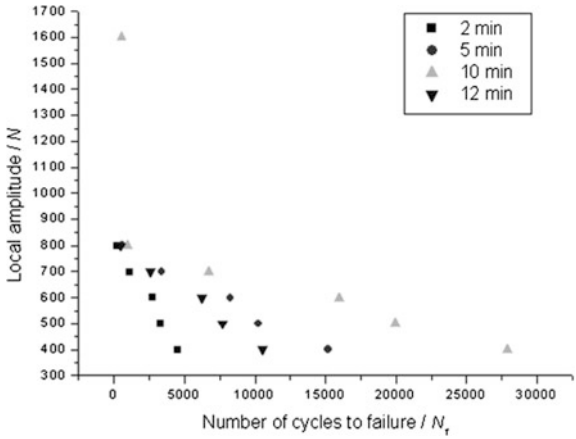


Fig. 1 Fatigue life under various remelting time

Table 1 Low cycle fatigue performance date of 10 min remelting sample

Δεt/2/%	Δεe/2/%	Δεp/2/%	Δσ/2/MPa	2Nf/cycle
0.242844	0.232291	0.010553	2.91335	18
0.592,565	0.54561	0.046955	5.67146	54
0.682815	0.63223	0.050585	6.5787	72
0.839627	0.778025	0.061602	8.2452	90
1.028533	0.931545	0.096988	9.3368	126
1.084999	0.975175	0.109824	9.89958	144
1.183373	1.059305	0.124068	11.0494	162
1.245532	1.112905	0.132627	11.77225	180
1.336937	1.181075	0.155862	12.7426	216
1.364843	1.259105	0.105738	13.0264	234
1.40828	1.29773	0.11055	13.4242	252
1.458905	1.28922	0.169685	13.9094	288
1.488346	1.32581	0.162536	13.8245	306
1.504563	1.431635	0.072928	14.0064	324
1.546189	1.38902	0.157169	14.3193	360
1.556093	1.38674	0.169353	14.42365	378
1.574781	1.50536	0.069421	14.59345	396
1.598468	1.52838	0.070088	14.68565	432
1.608252	1.469495	0.138757	14.82875	450
1.613904	1.53514	0.078764	14.9088	468
1.630065	1.448925	0.18114	15.06165	504
1.637967	1.55751	0.080457	15.12955	522
1.644062	1.56704	0.077022	15.0592	540
1.652782	1.527415	0.125367	15.19505	576
1.65853	1.555165	0.103365	15.24595	594
1.661439	1.498305	0.163134	15.3115	612
1.668311	1.560665	0.107646	15.33575	648
1.672344	1.560255	0.112089	15.3891	666
1.674748	1.60165	0.073098	15.33815	684
1.678095	1.60846	0.069635	15.4085	720
1.681282	1.61242	0.068862	15.343	738
1.685564	1.58195	0.103614	15.40855	774
1.688409	1.6175	0.070909	15.42305	792

In order to express low cycle fatigue performance of material, the famous elastic-plastic stress-strain relationship was proposed by Remberg-Osgood.

$$\frac{\Delta\varepsilon_t}{2} = \frac{\Delta\varepsilon_e}{2} + \frac{\Delta\varepsilon_p}{2} = \frac{\Delta\sigma}{2E} + \left(\frac{\Delta\sigma}{2k}\right)^{1/n'} \tag{3}$$

where, $\Delta\varepsilon_t$ is total strain amplitude, $\Delta\varepsilon_e$ is elastic strain amplitude, $\Delta\varepsilon_p$ is plastic strain amplitude, E is elastic modulus, $\Delta\sigma$ is stress amplitude, K is hardening constant and n' is hardening exponent.

Fitting parameter E.

From Eq. 3, $\Delta\varepsilon_e = \frac{\Delta\sigma}{E}$;

Let $E^{'} = \frac{1}{E}$, than $\Delta\varepsilon_e = E\Delta\sigma$;

Let $Q = \sum_{i=1}^{n} \left(E^{'}\Delta\sigma_I - \Delta\varepsilon_{ei}\right)^2$, than $\frac{\mathrm{d}Q}{\mathrm{d}E^{'}} = \sum_{i=1}^{n} 2(E^{'}\Delta\sigma_I - \Delta\varepsilon_{ei})\Delta\sigma_i$;

Let $\frac{\mathrm{d}Q}{\mathrm{d}E^{'}} = 0$, than $E^{'}\sum_{i=1}^{n}(\Delta\sigma_i)^2 - \sum_{i=1}^{n}\Delta\varepsilon_{ei}\Delta\sigma_i = 0$;

Hence, $E^{'} = \frac{\sum_{i=1}^{n}\Delta\varepsilon_{ei}\Delta\sigma_i}{\sum_{i=1}^{n}(\Delta\sigma_i)^2}$;

And $E = \frac{\sum_{i=1}^{n}(\Delta\sigma_i)^2}{\sum_{i=1}^{n}\Delta\varepsilon_{ei}\Delta\sigma_i}$ (4)

Fitting parameter K and $n^{'}$.

From Eq. 3: $\frac{\Delta\varepsilon_p}{2} = (\frac{\Delta\sigma}{2K})^{\frac{1}{n'}}$, $\ln\left(\frac{\Delta\varepsilon_p}{2}\right) = \frac{1}{n'}\ln(\frac{\Delta\sigma}{2K})$;

Let $\varepsilon = \ln(\frac{\Delta\varepsilon_p}{2})$, than $\varepsilon = \frac{1}{n'}(ln\Delta\sigma - ln2K)$ and $n^{'} = \frac{1}{\varepsilon}(\ln\Delta\sigma - \ln 2K)$ (5)

Let $Q = \sum_{i=1}^{n}\left[\frac{1}{n'}(ln\Delta\sigma_i - ln2K) - \varepsilon_i\right]^2$, as $\varepsilon_i = \ln(\frac{\Delta\varepsilon_{pi}}{2})$;

With solving partial derivative and make $\frac{\partial Q}{\partial n'} = 0$ and $\frac{\partial Q}{\partial K} = 0$:

$$\begin{cases} n\mathrm{x}^2 + D\mathrm{n}^{'}x - 2Bx - C\mathrm{n}^{'} + A = 0 \\ nx - D\mathrm{n}^{'} - B = 0 \end{cases} \tag{6}$$

$$A = \sum_{i=1}^{n}(ln\Delta\sigma_i)^2, B = \sum_{i=1}^{n} ln\Delta\sigma_i, C = \sum_{i=1}^{n} ln\Delta\sigma_i, D = \sum_{i=1}^{n}\varepsilon_i = \sum_{i=1}^{n}\ln(\frac{\Delta\varepsilon_{pi}}{2}) \text{ and } x = \ln 2K;$$

Equation 6 is a quadratic equation which could be solved by mathematic software such as Maple. From $x = ln2K$, $K = \frac{1}{2}e^x$ could be achieved.

Hence, E, K, n′ were fitted by least square method.

With applying the experimental results in Table 1 to Eq. 2: E = 9.986110536 GPa, K = 114.1876634 MPa, n′ = 0.932953886. The Δa could be calculated as listed in Table 2.

4.4 Least Square Method to Fit Parameters in 3USE-Fatigue Life Prediction Model

Fitting parameter m and C

In the 3USE fatigue life prediction model $N_f(\Delta a - \Delta a_0)^m = C$, the fatigue limitation could always be achieved by experiment. The below fitting process of m and C was carried out by considering the Δa as known.

Table 2 Low cycle fatigue performance date disposed by Remberg-Osgood equation

Δa/(M J • m^3)	Nf/cycle	Δa/(M J • m^3)	Nf/cycle
1.575244	9	52.31455	216
8.109604	27	58.42255	225
10.7188	36	54.24401	234
16.49347	45	63.32485	252
23.9269	63	55.90884	261
27.14948	72	55.56279	270
33.29723	81	60.15837	288
37.46427	90	58.78657	297
44.42518	108	63.89892	306
42.73802	117	59.77506	324
45.5461	126	60.46355	333
52.88832	144	57.2196	342
52.86433	153	57.30704	360
47.47168	162	57.0994	369
56.01236	180	60.26665	387
57.62233	189	57.78172	396
51.24402	198		

As $\Delta a_0 > \Delta a$, make a $= \Delta a_0 - \Delta a$, than $n_f a^m = C$;
Apply $\ln(N_f) = \ln C - \mathrm{m} \ln a$ and $\varepsilon = \ln(N_f)$, than $\varepsilon = \ln C - m \ln a$
Let $\mathrm{Q} = \sum_{i=1}^{n} [\ln C - m \ln a_i - \varepsilon_i]^2$ (1)
As $a_i = \Delta a_0 - \Delta a_i$, where Δa_i could get from Eq. 2 and $\varepsilon_i - \ln N_{fi}$
With solving partial derivative and make $\frac{\partial \mathrm{Q}}{\partial \mathrm{C}} = 0, \frac{\partial Q}{\partial m} = 0$

$$\begin{cases} nx - Am - B = 0 \\ Ax - Fm + D = 0 \end{cases} \tag{2}$$

As $A = \sum_{n=1}^{n} \ln a_i, B = \sum_{n=1}^{n} \varepsilon_i, F = \sum_{n=1}^{n} \ln a_i^2, D = \sum_{n=1}^{n} \varepsilon_i \ln a_i \quad and\, x = \ln C$

Equation 8 was linear equations related to x and m which could be solved easily. And from $\mathrm{x} = \ln C$, $\mathrm{C} = e^x$ could be achieved.

Construct fatigue life prediction model

With using the experimental date in Table 2, Δa_0 = 52 M J • m^3, m = 0.045827313,C = 168.3815495 were determined in the case of 10 min. Hence, the 3USE-fatigue life prediction model under the situation of 10 min remelting treatment is conducted as below:

$$N_f(\Delta a - 52)^{0.045827313} = 168.3815495$$

5 Results

1. A test on combined bending and torsion fatigue life of flame thermal spray components was carried out in this study. Hence, test date of specimen under best remelting time situation was achieved.
2. With applying the least square method, the parameters of elastic-plastic stress-strain proposed by Remberg-Osgood were fitted. Hence, parameters related to Δa (fatigue parameter amplitude) in 3USE method was achieved: Δa was calculated by E, K, n' which were fitted in algebraic expression and came from low cycle fatigue test.
3. With applying the least square method, m and C in 3USE method were fitted and calculated by fatigue test. Hence, the prediction model with using 3USE was constructed.
4. As the uniform 3USE method could employ both high cycle fatigue and low cycle fatigue, the fatigue life of corresponding thermal spraying components could be predicted with using only the $\Delta\varepsilon e$, $\Delta\varepsilon p$, $\Delta\sigma$ and Nf which could get from the normal fatigue test.

References

1. Guo-Qing Z, Cheng-Tao W, Bin-Shi X (2011) On assessment of several fatigue life prediction methods. J Mech Strength 33(3):469–474
2. Coffin LF (1970) The effect of frequency on high temperature low-cycle fatigue. In: Proceedings of air force conference on fracture and fatigue of aircraft structures, AFDL-TR-70-144, pp 301–312
3. Ostergren WJ (1976) A damage function and associated failure equation for predicting hold time and frequency effects in elevated temperature low cycle fatigue. J Test Eval 4(5):327–339
4. Manson SS, Halford GR, Hirschberg MH (1971) Creep-fatigue analysis by strain-range partitioning. In: Design for elevated temperature environment, ASME, New York, pp 12–28
5. Coffin LF (1976) The concept of frequency separation in life prediction for time-dependent fatigue. In: Symposium on creep-fatigue interaction, ASME-MPC-3, New York, pp 349–363
6. He J, Duan Z, Ning Y et al (1983) Strain energy partitioning and its application to GH33A nickel-base superalloy and 1Cr18Ni9Ti stainless steel. In: Woodford DA, Whitehead JR (eds) In advance in life prediction method. ASME, New York, pp 27–32
7. Miller MP, McDowell DL, Ochmker LF et al (1993) A life prediction model for thermomechanical fatigue based on microcrack propagation. In: Sehitogla H (ed) Thermomechanical fatigue behavior of materials. ASTM STP 1186, American Society for Testing and Materials, Philadelphia, pp 35–49
8. Xu C, Zhang G-D, Su B (2007) A method based on energy and three-parameter power function for low cycle fatigue. J Mater Eng 8:65–68
9. Zhao Z-P, Li X-Y, Li Y-T et al (2012) Analysis of fatigue properties for plunger with different post-fused thermal spray. Trans Mater Heat Treat 33(S1):92–95
10. Zhao Z-P, Li X-Y, Duan H-Y, Li Y-T (2012) Influence of coating thickness on tension and compression fatigue performance of flame thermal spraying components. Trans Mater Heat Treatment 33(S2):139–144

11. Zhao Z-P, Li Y-T, Li X-Y (2013) Effects of remelting time on fatigue life and wear performance of thermal spray welding components. Transactions of Materials and Heat Treatment 34(7):169–174
12. Zhao Z-Ping, Li X-Y, Li Y-T et al (2012) Effect of remelting processing on fatigue properties of Ni based PM alloy parts with thermal spraying coating. Powder Metall Techno (1):3–7

Design of Poultry Egg Quality Detection System Based on LABVIEW and PLC

Yan Liu and Rui-Jie Kong

Abstract The ability to measure the size of the eggs, the ability to choose between the right and the bad eggs, and the competitiveness of the eggs in the market, also affect the economic benefits of the enterprises. In this paper, the technology of PLC and LABVIEW is used to design a system for the identification of the dirty spots and cracks and quality classification The control scheme and control flow chart are given, and the communication connection and data exchange contents of LABVIEW and PLC are established. the method of gray-scale, binarization and edge tracking of the image collected by LABVIEW software is analyzed, and the characteristics of the extraction of dirty spots and crack images are obtained. This method has realized the fast speed and intelligence of poultry egg quality detection, and has practical application value.

Keywords LABVIEW · Quality detection · Dirty spots and cracks
Image processing · Quality classification

1 Introduction

Eggs are a kind of animal food, rich in high protein, low fat, a variety of vitamins and minerals, is the general diet of universal consumer goods. When the farm output of eggs, how to quickly and accurately detect the size of eggs, to distinguish between the advantages and disadvantages of eggs, a great impact on the egg packaging, transportation, custody and acquisitions. But also related to the competitiveness of eggs in the market, affecting the economic efficiency of enterprises [1]. The use of PLC and LABVIEW technology integration, the design process on the egg dirty spots and cracks in the identification, for quality grading, can be efficient and fast to achieve the quality of egg testing and classification. LABVIEW

Y. Liu (✉) · R.-J. Kong
School of Electrical and Automotive Engineering, Changshu Institute of Technology, Changshu, China
e-mail: lyan@cslg.edu.cn

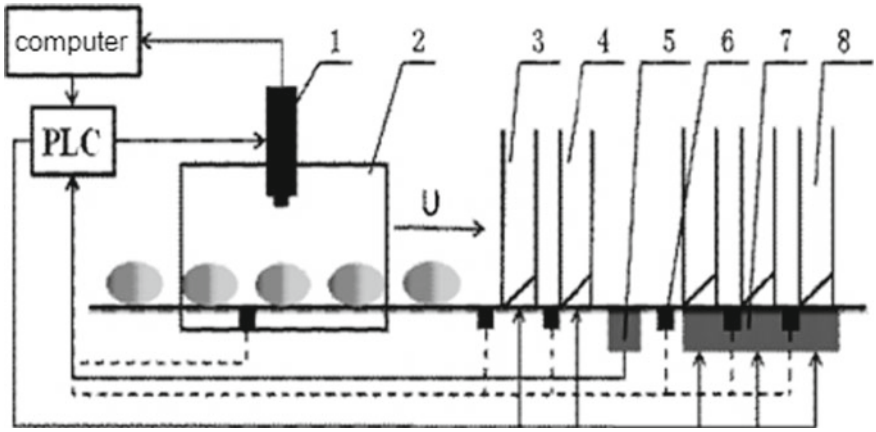
K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_31

is used to complete the collection and processing of egg images, the thresholds of the collected images are segmented, and the identification of the dirty spots and cracks is completed. The PLC controller is used to complete the sorting and quality classification of the eggs. To achieve rapid detection of egg quality, intelligent degree. This can reduce the damage rate in the process of detecting the classification. Reduce the staff's work pressure, improve their work efficiency. This study has practical value and significance [2, 3].

2 The Structure and Control Requirements of the Egg Detection System

2.1 The Structure of the Egg Detection System

The structure of the egg detection system shown in Fig. 1, the system consists of camera, light box, remove the dirty spot channel, remove the crack channel, weighing module, sensor, weight grading channel and other components. The camera completes the collection of egg images. A number of photoelectric sensors placed in different locations, used to complete the order of the egg running the process of resolution, to facilitate the realization of dirty eggs, crack eggs and different quality eggs separated selection. The weighing module is used to detect the quality of eggs. When the eggs are moving in the longitudinal direction of the conveyor chain, they are also rotated 360° on the pallet mechanism, which ensures that all areas on the egg shell are collected. When the acquisition area reaches the camera real-time acquisition of dynamic images. After the acquired image is



1. Camera; 2. Light box; 3. Remove dirty spot channel; 4. Eliminate crack channel;5. The weighing module; 6. The sensor;7.Sizing actuators;
8. Weight classification channel

Fig. 1 Egg testing and hierarchical control system. structure

processed, the information of the dirty spots egg and the cracked egg is obtained. If It is the dirty spot egg, It will be removed, the eggs without dirty spots into the next crack recognition program, if a crack is present there will be removed too, for eggs neither dirty spots nor cracks can into the next step which is quality grading control.

2.2 The Design of Control System

Based on PLC and LABVIEW technology, the image recognition and control system based on PLC and LABVIEW technology is used to collect and process the eggs image by LABVIEW software. The image features of dirty and cracked eggs can be identified, and the recognition signals are passed into PLC and PLC by LABVIEW. Signal characteristics after the corresponding control of the actuator solenoid valve to remove the dirty spots and cracks in the eggs, the non-dirty points and cracks in the eggs were separately controlled, so that the sorting agencies to complete the quality grading. The control process is as follows:

1. Place the photoelectric sensor at the front of the camera to detect whether the eggs are arriving. When the eggs arrived, the photoelectric sensor will send a signal to the camera began to work, the use of LABVIEW to achieve the collection of egg images.
2. When the eggs reach the dirty point channel, the photoelectric sensor will be triggered, the use of LABVIEW image processing functions, the image will be grayed out, binarized, remove the boundary particles, boundary tracking, seeking boundary area processing, extraction Graphic features about dirty spots and cracks. LABVIEW will send the result image of dirty or crack recognitioninto the PLC variable memory (VB0, VB2) [4, 5].
3. According to the recognition result of LABVIEW image, the removal of egg dough and crack is completed under PLC control. If the eggs have dirty spots or cracks, then remove the dirty or crack the solenoid valve. For non-dirty and non-cracked eggs, the load cell converts the result into an analog voltage signal and enters the PLC for A/D conversion. When the eggs have entered the corresponding level of the channel, the photoelectric sensor is triggered, Which controls the handling of different solenoid valves for mass grading [6, 7]. The egg quality grading requirements are shown in Table 1. Detection and control process shown in Fig. 2.

Table 1 Egg weight grading

Weight	Eggs level
≥ 68 g & <78 g	LV XXL
≥ 60 g & <68 g	LV XL
≥ 53 g & <60 g	LV L
≥ 48 g & <53 g	LV M
43 g & 48 g	LV S
<43 g & ≥ 78 g	Defective products

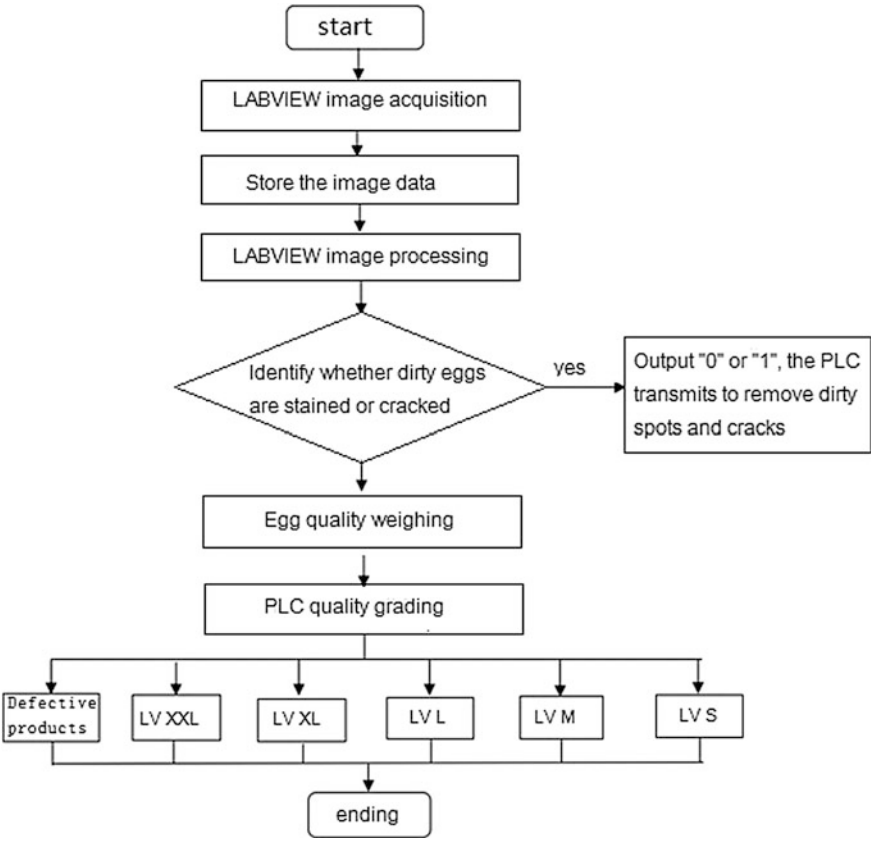


Fig. 2 Detection and control flow chart

2.3 A Communication Connection with the PLC and LABVIEW

In order to establish communication between Siemens PLC and NI LABVIEW, we will use OPC server software based on Siemens. Download and install PC Access software and DCS toolkit, so that S7-200 and PC access through PC/PPI cable to take PPI protocol, PC access and LABVIEW go OPC protocol, that is, through the 'My Computer-New-I/O Server-S7-200server, create I/O server'. After the communication is built, the data link with PLC and LABVIEW will also be established.

According to inspect and control the eggs in the PC access with LABVIEW need to exchange the data content, as shown in Fig. 3. Set up bind variables, namely SQ1 and photoelectric sensor is installed in a PC access project M0.2 binding; Dirty spot identification of photoelectric sensor for SQ2 with PC access to establish project M0.3 binding; Crack identification of SQ3 and photoelectric sensor is installed in a PC access project M0.3 binding; Classification of the photoelectric sensor SQ4

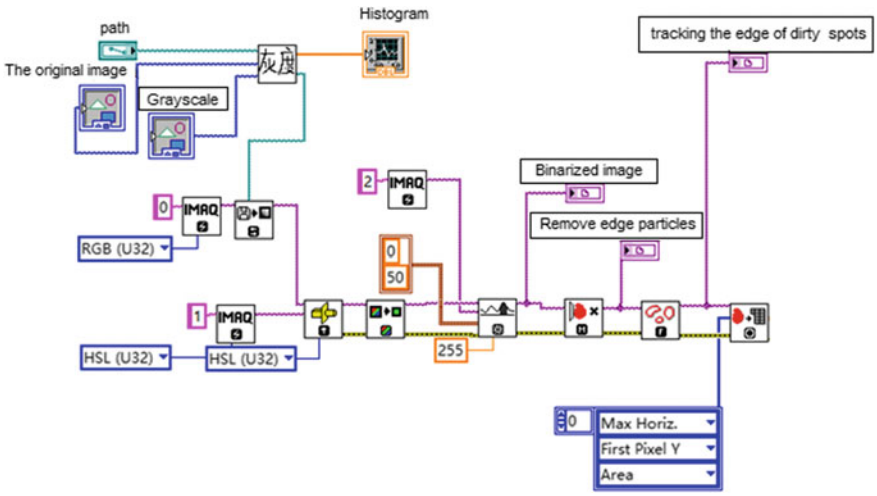


Fig. 3 Dirty spots recognition program

M0.5 binding with PC access to establish project; Weight level XXL solenoid valve YV3 with PC access to establish project Q0.4 binding, and so on. After such data link, through LABVIEW in the PLC to control the outcome of the image processing on the corresponding switch, convenient operation personnel real-time monitoring.

3 The Image Processing of Dirty Spots and Cracks Egg

3.1 The Image of Egg Collection Process

LABVIEW has strong advantages in image processing, motion control, data analysis and chart display. This design USES LABVIEW to complete the collection and processing of egg images. When the eggs arrive at the detection range of the photoelectric sensor SQ1, the camera will be triggered and the image of the collected eggs will be activated and transferred to the computer. The program design idea is as follows:

The design process of LABVIEW is as follows: the Camera is connected to the IMAQdx Open Camera VI control to control the opening and closing of the Camera, and then to the IMAQdx Configure Grab VI control to collect the scanned images. Also connect the camera to the IMAQ Create VI letter control to Create the image storage path. Finally, the first two steps are connected to the input points corresponding to the IMAQdx Grab VI function, and the image obtained from the last photo is collected. In order to ensure that program can collect multiple images, it is necessary to add loops and conditions on the fringes of the program structure, ensure that only when the photoelectric sensor to trigger Image acquisition processing, and displayed on the Image on the front panel.

3.2 Image Grayscale

Image Grayscale Transformation The image data cluster is converted to a two-dimensional array through the Restore Pixel Chart (VI) control. After the image is converted to a grayscale image, we can create an array of one-dimensional 256 elements and initialize it to zero. And then call the 'replace array subset' and 'index array' control to the image pixel gray value for the index, draw the gray histogram. Image grayscale is a means of converting a color image into a grayscale image, with the aim of reducing the amount of information on the image. The description of the grayscale image, like the color image, still reflects the distribution and characteristics of the overall and local chroma, brightness level of the image.

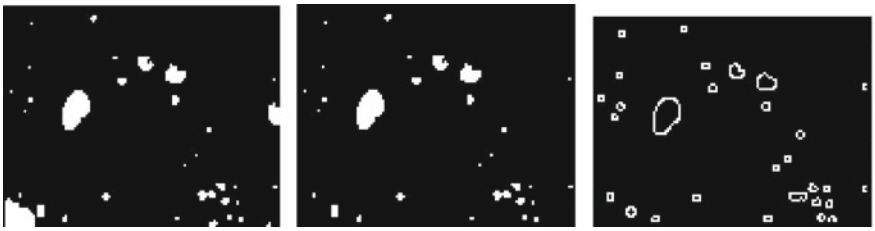
3.3 Dirty Spots Feature Extraction

For binocular images that have been grayscale, binarization is performed to further determine their thresholds. The specific approach is to observe the gray histogram, find the point in the curve, the first appeared, the amplitude of the larger fluctuations in the point, that is, 'valley point', then the gray value is the threshold. First, the IMAQ Read File VI control reads the image file and converts the current image type to the specified image type through the IMAQ Cast Image VI control, reads the birdFs dirty point by the gray histogram Threshold.

Using the IMAQ Threshold VI control for binarization, we need to collect the stained image several times and then perform the grayscale processing. Then, observe the histogram of the gray scale, and get the threshold range of the dirty point, that is, the minimum gray value of the range port of the IMAQ Threshold VI control is set to 0, the maximum is 50, and the dirty point of the egg is from the whole egg image Split out. The IMAQ Reject Border VI control fills the edge tiny particles, and the IMAQ Canny Edge Detection VI controls the edge tracking of the dirty points. This completes the extraction of dirty points. The dirty point recognition function is shown in Fig. The dirty point extraction image is shown in Fig. 4.

3.4 Cracks Feature Extraction

As with the dirty point extraction procedure, the image file after the gray scale is first read by the IMAQ Read File VI control. The current image type is converted to the specified image type by the IMAQ Cast Image VI control, and the threshold of the egg crack is read by the gray histogram. We binarize the IMAQ Threshold VI control, repeat the acquisition of the crack image several times, and then perform the gray scale processing to observe the gray histogram, which can roughly approximate the threshold of about 170. That is, the minimum gray value of the



(a) Dirty spots binarized image (b) The particle image of removing dirty spots edge (c) the image of the edge of dirty spots

Fig. 4 Extract dirty spots image

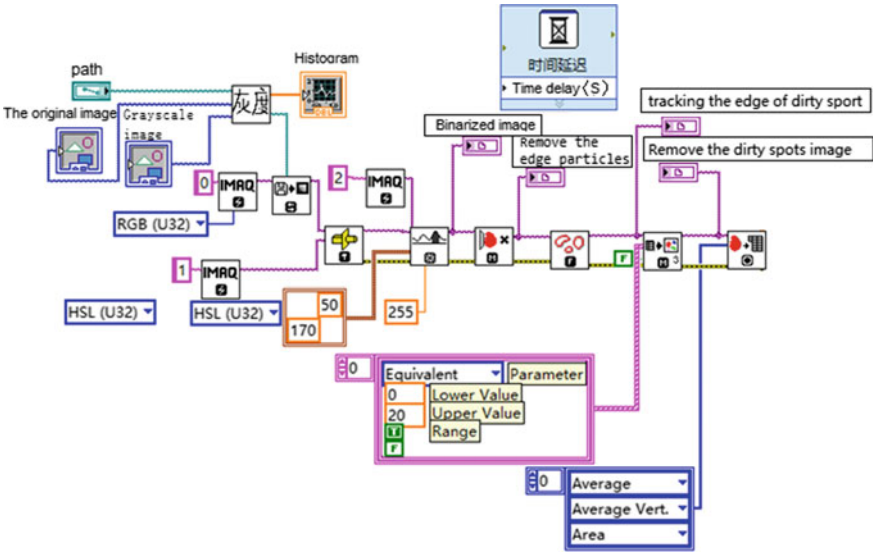
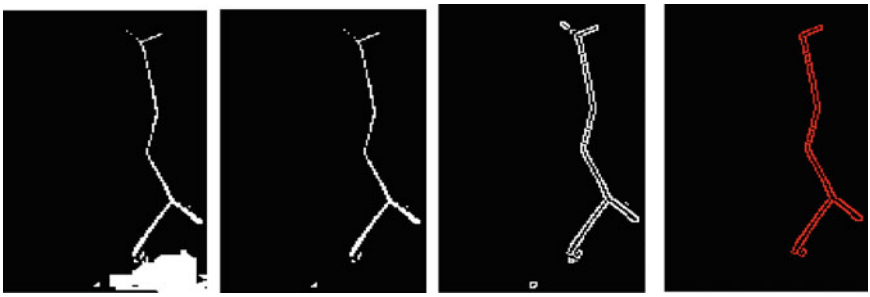


Fig. 5 The crack recognition program

range port of the IMAQ Threshold VI control is set to 50 and the maximum is set to 170, and the cracked eggs are separated from the image according to the threshold size. In particular, because the dirty point of the gray value range is relatively wide, there may be individual dirty point eggs, the gray value has reached 170, with the crack and split out. Let the IMAQ Reject Border VI control fill the edge tiny particles so that the IMAQ Canny Edge Detection VI controls the edges of the split image. Because it is also possible to extract some dirty points at this time, so IMAQ Particle Filter 3 VI control must be added to remove the dirty dot image. Finally, the extracted image is completely cracked. Crack extraction program block diagram shown in Fig. 5, extract the image shown in Fig. 6.



(a) Binarized image (b) Remove the image of the edge particles of the edge (c) Track the image remove the dirty spots (d) the image after

Fig. 6 Crack extraction image

3.5 Identification Dirty Spots and Cracks

The IMAQ Particle Analysis VI is functionally analyzed based on the dirty and cracked images, and the area of each dirty or crack image is calculated. If the presence area is greater than 0, there is a dirty point or crack. At the same time, the dirty or cracked display controls output '1' (otherwise the output is '0'). At the same time LABVIEW will dirty point or crack display control output number '1' into the PLC, PLC according to the received information to control the dirty point or crack egg solenoid valve action. Finish dirty or cracked eggs. The dirty point or crack judgment procedure is shown in Fig. 7.

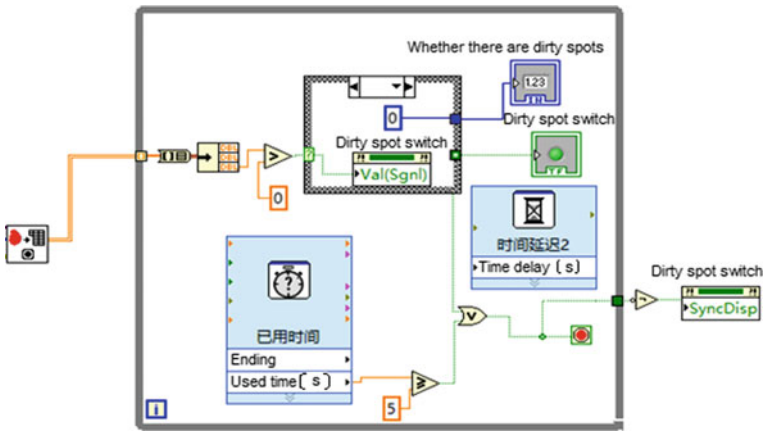


Fig. 7 FIG program dirty spots or cracks on Analyzing

4 Analysis of Test Results

According to the procedures described in this article, the composition, detection and processing procedures. The eggs with dirty spots and cracks were identified. Get a satisfactory test results.

(1) Detection of dirty eggs. When the eggs enter the dirty channel. Trigger dirty point to identify photoelectric sensor. LABVIEW performs dirty dot identification according to the procedure described above and the image processing method. If it is dirty, the corresponding indicator on the front panel of the LABVIEW light is on and the dirty point control is displayed as 1.
(2) Detection of cracked eggs. When the eggs enter the crack channel. A photoelectric sensor that triggers a crack recognition. LABVIEW starts running the crack recognition program. When the detection of eggs in the presence of cracks, the corresponding front panel lights light. The crack display control is displayed as 1. When the eggs are clean and no cracks, LABVIEW front lights are not bright. The results of the LABVIEW crack recognition front panel are shown in Fig. 8.
(3) Quality grading. Because the PLC and LABVIEW established a communication connection between. We can observe the sorting status of the egg quality of the PLC from LABVIEW. When the quality of the PLC on the solenoid valve trigger, it will cause LABVIEW front panel to run. When the eggs into a certain level of the channel that is triggered by the channel of the sensor, the egg is the level of quality. Figure 8 for the quality rating debugging interface. The figure shows that the LABVIEW front panel M-level indicator lights up when the PLC drives the M-channel switch.

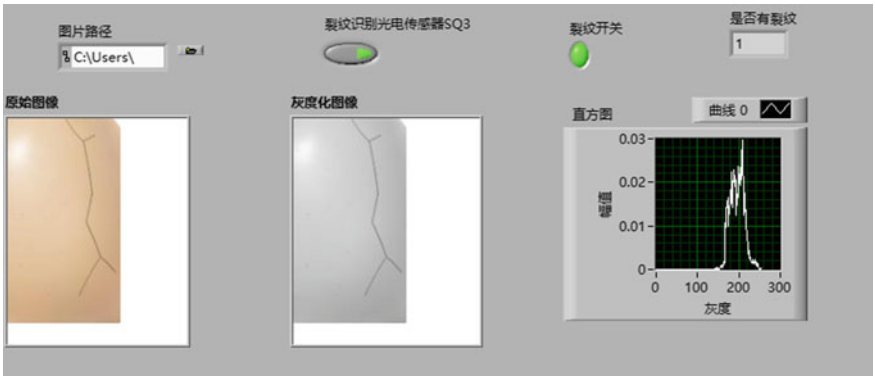


Fig. 8 Identification of crack shows

5 Conclusion

This design uses LABVIEW and PLC technology combined. Identify the egg dough and crack. And the quality of grapes without cracks and cracks in the grading. LABVIEW as an important programming language, provides a convenient way to achieve instrument programming and data acquisition. So that we can use LABVIEW to carry out the identification of egg dough and crack. The identification of egg dough and crack is the core of this design. In the collection of pictures must be selected the appropriate light source, and light background images can not have shadows. Let the image be collected close to the object's true color. To ensure that the extraction of dirty points and cracks gray and the background of the gray between the existence of a certain difference. Meet the requirements on the gray image threshold identification and processing. LABVIEW software to implement the image gray, binarization, edge tracking and other processing to save a lot of complex algorithms. It achieves the identification of dirty spots and cracks. This design shortens the time for egg quality testing and grading. Reducing the damage rate of eggs during the detection and classification process. Improve work efficiency. It has practical application value.

References

1. Xin L, Zhi-Hong Y (2009) Application of dynamic image acquisition in egg quality testing. J Agricu Mechanization Res 9:22–45
2. Liang W, Zhi-Hong Y, Lu W, Zhi-Hui W (2013) Design of egg automatic detection grading and packaging device. J Agricu Mechanization Res 3:53–61
3. Zhi-Hui W, Cao Y, Lu W, Lang W (2013) Development on the plc control system for the egg quality inspection and grading[J]. J Inner Mongolia Agricu Uni 4:123–126
4. Su L (2013) Research on the key technology of egg quality online intelligent detection. Jiangsu University
5. Wei-Da S, Zhao-Peng H, Xu-Ming L (2013) Research status of fresh egg grading and quality grading technology. National Egg Eng Techno Res Center 8:4–11
6. Yan Z-D, Sun L-, Hu C-G (2012) Temperature control system based on LABVIEW. Instrument Tech Sensor (9):13–14
7. Xiao-bo LI (2013) Vehicle Water Temperature Sensor Test System Based on LabVIEW. Instrument Tech Sensor 1:59–67

Thermal Deformation of a Vertical Plate According to Various Plastic Deformation Regions

Teng Yao, Lanzhong Guo and Chao Wang

Abstract Thermal straightening is a widely used process for dealing with undesired distortion. Flame heating and induction heating are the traditional heating processes used in industry. However, these two heating processes have an insignificant effect on vertical steel plates. This study compared the thermal deformation caused by induction heating with the thermal deformation caused by arc heating and laser beam heating which have greater heat input. The thermal deformations induced by these three processes were simulated by established heat source models, and simulated results were compared with experimental results to confirm the model's accuracy. With investigating the mechanism of thermal deformation with the finite element method it was found that the thermal deformation have a significant relationship with the amount of central zone plastic deformation.

Keywords Thermal straightening · Thermal deformation · Line heating
Arc heating · Induction heating · Laser beam heating

1 Introduction

In normal industries, thermal straightening can be applied to a welded structure to reduce welding deformation. It is recommended that the thermal straightening limit the maximum temperature of steel to below the temperature of solid state phase transformation.

T. Yao
School of Mechatronic and Electrical Engineering, China University of Mining and Technology, Xuzhou, China
e-mail: 970722509@qq.com

L. Guo · C. Wang (✉)
School of Mechatronic Engineering, Changshu Institution of Technology, Suzhou, China
e-mail: cslgcw@cslg.edu.cn

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_32

Since the temperature is limited, the process has little effect on small structures, especially vertical plates. Therefore, arc heating and laser beam heating processes were induced in this study. The thermal deformation caused by these processes was compared to the thermal deformation caused by induction heating.

The finite element method has become a powerful tool for a variety of scientific and engineering problems since its development in the 1940s. In recent years, Park et al. [1] proposed a new control method for longitudinal bending in T-Joint beams under induction heating with the help of three-dimension thermal elastic plastic analysis.

Multiple numeric models using FEM have been developed for predicting thermal deformation in the arc welding and laser beam welding [2]. Laser beam heating produces a narrower heat-affected zone, which results in lower distortions. Tsirkas et al. [3] used numerical simulation to study the effects of laser beam heating parameters on thermal deformation, and found the most important parameters of laser beam heating to be laser power and laser speed. Moraitis et al. [4] used a FEM to predict the residual stresses and distortions caused by laser beam heating. With this keyhole model developed as a heat source, the difficult and expensive experimental tests required for keyhole characterization can be greatly reduced. Using the same power scale as arc heating, laser beam heating produced a smaller amount of thermal deformation.

2 Materials and Methods

40Cr steel is widely used for structural parts because of its strength, toughness, and wear resistance. Specimens were laser cut to the dimensions of 100 $\times$ 40 $\times$ 7 mm. Because of its high stiffness, the deformation induced in this material is much lower than that of other metals.

2.1 Induction Heating

Induction heating is traditional heating process in industry. Its advantages include speed, easy handling, and automation capability. In this process, heat is formed by induced eddy currents. The coil shape and heating position are specially designed to meet the desired heating objective. Two circular heating coils to allow for heating both sides of a plate at the same time. The coils were fixed during the heating process, and the specimen had a relative velocity of 2 mm/s with the movement of foundation.

2.2 Arc Heating and Laser Beam Heating

In this study, both arc heating and laser beam heating were heated at center line of upside. The arc heating was performed by a tungsten inert gas welding machine

without filler added. The power was set at 2 kW with a constant velocity of 5 mm/s. The thermal distortion both before and after line heating was measured by a 3D coordinate measuring machine.

The advantages of laser beam heating include stability, flexibility, rapidity, and good quality. In particular, the power patterns used in laser beam heating can be adjusted freely by engineers. The experiment in this study was carried out using disk laser equipment. The laser beam has a diameter of 0.6 mm and a focal length of 450 mm. A similar 2 kW was selected as heating power.

3 Numerical Analysis

In this study, FEM was performed to calculate the thermal history and thermal deformation for a wide range of cases. The experimental cases described above were used to verify the numerical model.

A 3D solid elastic plastic model was performed in all numerical simulations. In a full consideration, the total strain $\boldsymbol{\varepsilon}$ can be decomposed into the sum of the thermal strain $\boldsymbol{\varepsilon^T}$, the elastic strain $\boldsymbol{\varepsilon^E}$, the plastic strain $\boldsymbol{\varepsilon^P}$, the creep strain $\boldsymbol{\varepsilon^C}$ and the solid-state phase transformation strain $\boldsymbol{\varepsilon^{phase}}$ as shown in Eq. (1). However, because the time period above the creep temperature was very short, and because of a nature of low carbon contained, the creep strain and phase transformation strain were ignored in this study [5].

$$\boldsymbol{\varepsilon^{Total}} = \boldsymbol{\varepsilon^T} + \boldsymbol{\varepsilon^E} + \boldsymbol{\varepsilon^P} + \boldsymbol{\varepsilon^C} + \varepsilon^{Phase} \tag{1}$$

Numerical simulation of heat transfer considers two kinds of heat: heat that moves from a metal surface to the surrounding fluid, and radiation. In the numerical simulation of the thermal process, both convection and radiation occur simultaneously. The total heat loss becomes the sum of Eqs. (2) and (3), where $\boldsymbol{\varepsilon}$ is the emissivity and $\boldsymbol{\sigma_0}$ is the Stefan-Boltzmann constant.

$$q_c \propto \beta_c (T - T_0) \tag{2}$$

$$q_r \propto \varepsilon\sigma_0 \left\{ (T + 273)^4 - (T_0 + 273)^4 \right\} \tag{3}$$

During the analysis, a simple but effective boundary condition was applied in order to prevent the rigid body motion. Three bottom-corner nodes were fixed in the z-direction, two nodes were fixed in the x-direction (heating direction), and one node was fixed in the y-direction. This boundary condition was used to avoid limiting expansion and shrinkage during heating and cooling cycles.

3.1 Induction Heating Modeling

Since induction heating has a large heat effect zone and low light penetration, a disk surface heat source with Gaussian distribution heat flux was modeled in the numerical simulation. An efficiency of 75% was used for the numerical simulation of induction heating. The heat source modeled by Eq. (4) is illustrated in Fig. 1:

$$q_s(x,y) = \frac{\alpha Q_s}{\pi r_s^2} \exp\left[-\frac{\alpha(x^2+y^2)}{r_s^2}\right] \tag{4}$$

3.2 Arc Heating Modeling

The tungsten inert gas arc heating process was simulated by applying a heat source of Goldak's double ellipsoidal distribution (Fig. 2) [6]. The heat flux is simulated by Eqs. (5) and (6):

For the front heat source:

$$\mathrm{q(x,y,z,t)} = \frac{6\sqrt{3} f_f Q}{abc_f \pi\sqrt{\pi}} \mathrm{e}^{-3x^2/a^2} \mathrm{e}^{-3y^2/b^2} \mathrm{e}^{-3(z-vt-z_0)^2/c_f^2} \tag{5}$$

For the rear heat source:

$$\mathrm{q(x,y,z,t)} = \frac{6\sqrt{3} f_r Q}{abc_r \pi\sqrt{\pi}} \mathrm{e}^{-3x^2/a^2} \mathrm{e}^{-3y^2/b^2} \mathrm{e}^{-3(z-vt-z_0)^2/c_r^2} \tag{6}$$

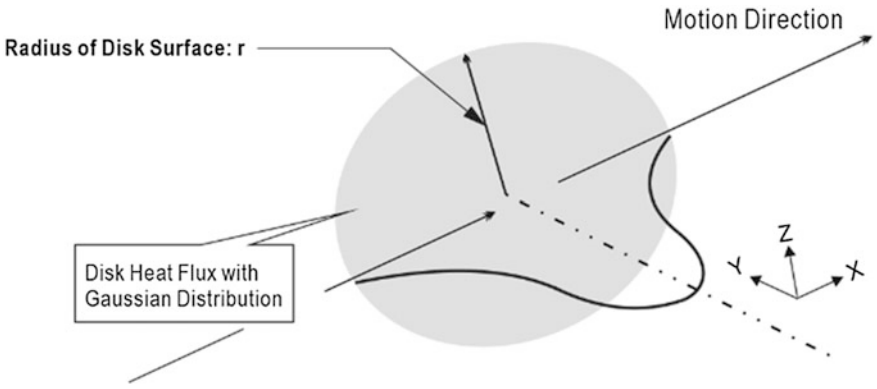


Fig. 1 Disk surface heat source model with Gaussian distribution heat flux

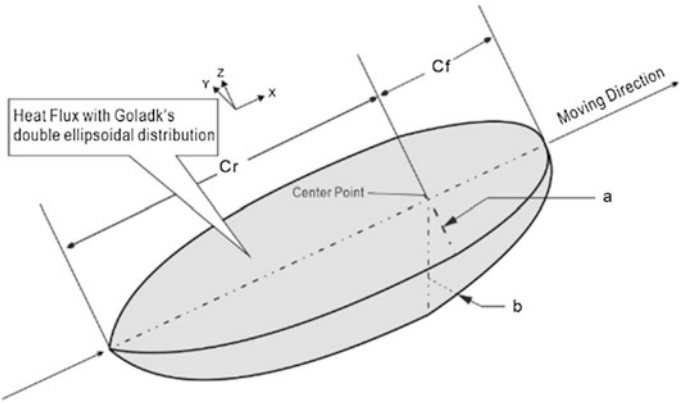


Fig. 2 Goldak's double ellipsoidal heat source model

Three cases with varying power were used in the simulation of arc heating. Different cases were set at 2.0 kW. The heating velocity was constantly 5 mm/s as that of experiments. From the shape of the bead by experiment, a:b: $c_f : c_r$ was given by 2.2 : 2.2 : 2.2 : 4.1. The simulation of heat source expressed by Eq. (7) was performed using values of $f_f = 0.695$ and $f_r = 1.305$. An efficiency of 70% was used for the numerical simulation of tungsten arc heating [7].

$$f_f = 2\frac{c_f}{c_f + c_r} = 2 - f_r \tag{7}$$

3.3 Laser Beam Heating Modeling

The advantage of laser beam heating is its locally high-density energy input. In the past decades, many authors have developed heat source models to predict keyhole style. In the current study, the heat source of the laser beam was modeled by a Gaussian surface distribution heat source and a cylinder volumetric heat source as shown in Fig. 3. The separated occupied power was expressed in the Eq. (8) with a partition distribution [8].

$$Q \times \eta = Q_{Gaussian\ sur.}(20\%) + Q_{cylinder\ vol.}(80\%) \tag{8}$$

Here η is the efficiency of laser beam heating, which was set at 95% in this study.

The heat flux from Gaussian surface distribution was modeled by Eq. (9) and the heat flux from cylinder volumetric distribution was expressed by Eq. (10).

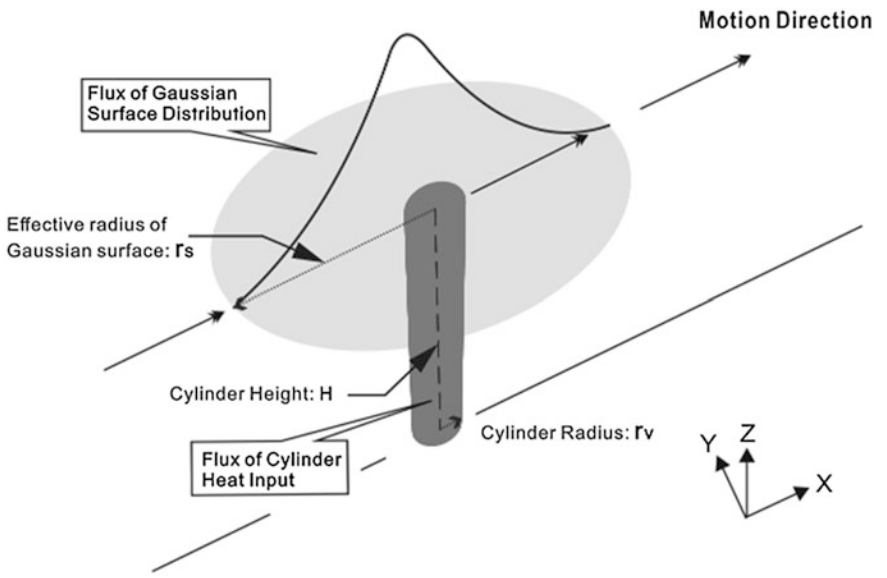


Fig. 3 Equivalent heat source for laser beam heating

$$q_s(x,y) = \frac{\alpha Q_s}{\pi r_s^2} \exp\left[-\frac{\alpha(x^2+y^2)}{r_s^2}\right] \tag{9}$$

Where, α is distribution parameter and r_s is effective radius of Gaussian surface distribution.

$$q_v(x,y) = \frac{6Q_v(H-\beta z)}{\pi r_v^2 H^2(2-\beta)} \exp\left[\frac{-3(x^2+y^2)}{r_v^2}\right] \tag{10}$$

Where β is the energy attenuation coefficient and H and r_v are the height and radius of the volumetric cylinder. Here, the heat flux parameters α, β, r_s, r_v, and H were set as 0.3, 0.15, 1.3, 0.6, and 2.0 mm in this study.

4 Results and Discussion

Thermal deformation is measured along the center line of the specimen's bottom. This distortion is generated by the bending moment caused by heating process, which has a direct relationship with heating input and heating position. A wide range of cases with different parameters was calculated in this study. The models developed in this study successfully predicted the thermal deformation caused by induction heating, arc heating and laser beam heating.

4.1 Thermal Deformation by Induction Heating

A total of three induction heating (IH) cases with differing heating heights were investigated by numerical simulation and experimentation. The insignificant thermal deformation predicted by induction heating was confirmed by experimental results (Fig. 4). Anyhow, the insignificant thermal deformation means the induction heating is not effective in this small steel plate deformation.

4.2 Thermal Deformation by Arc Heating

Based on previous work on the numerical simulation of tungsten inert gas welding, an arc heating experiment was carried out for the power of 2 kW. Figure 5 shows the compared results of arc heating (AH) experiment and simulation; these results are generally in agreement and the accuracy of the arc heating model was confirmed.

4.3 Thermal Deformation by Laser Beam Heating

The laser beam heating (LBH) experiments was compared with numerical simulation at Fig. 6. The peak thermal deformation produced by B0 was about −57 μm. Case 6 produced almost the same thermal deformation as the arc heating experimental case of 2 kW. It should be noted that the actual power input of laser beam heating was much larger than that of arc heating since the efficiency of laser beam is much higher.

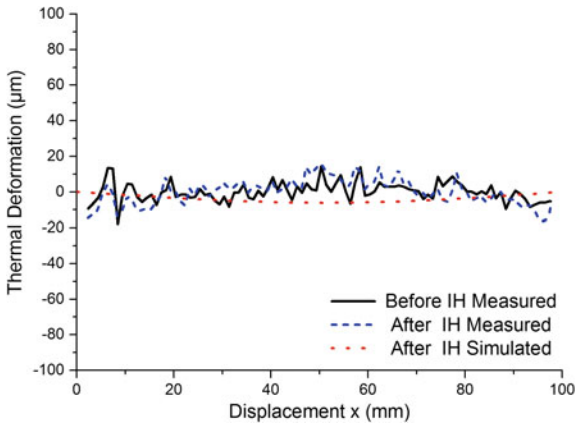


Fig. 4 Thermal deformation by induction heating

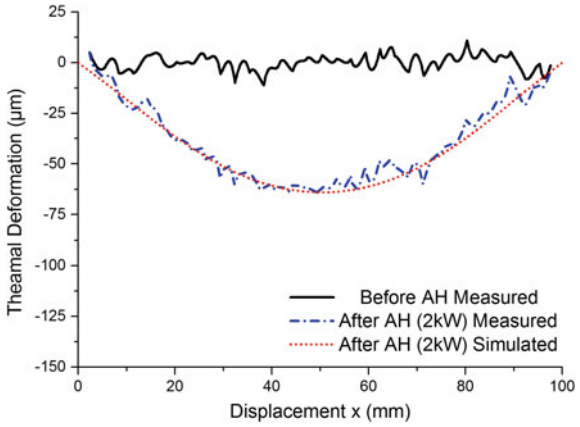


Fig. 5 Thermal deformation of 2 kW arc heating

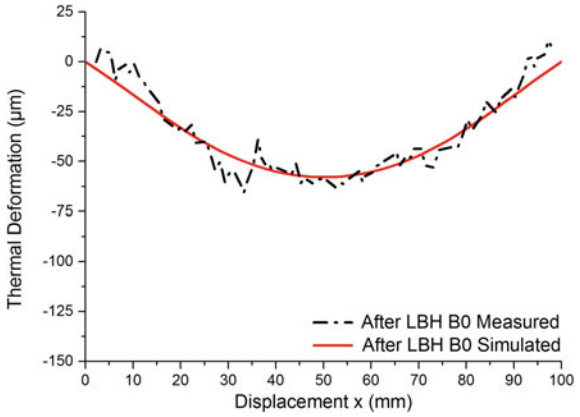


Fig. 6 Thermal deformation by laser beam heating B_0

This relationship could illustrate the magnitude of shrinkage by yield in the center zone. Hence, a much greater shrinkage force would be produced in the center zone, and the bending effect would be much larger after the plate's cooling to room temperature.

5 Conclusions

A basic study on the thermal deformation created by various heating processes was accomplished. Arc heating and laser beam heating process are both recommended to be used on small structures. The key points are summarized below:

1. Induction heating had an insignificant effect on the thermal deformation of small specimens. This may have been due to a wide heat-affected zone and a lower peak temperature.

2. Arc heating could be a convenient method for line heating. A power of 2 kW has a relatively large effect on the thermal deformation of small specimens. Thermal deformation ability will increase significantly with an improvement of power supply.
3. Since the plastic deformation region produced by laser beam heating is much concentered than that of arc heating, a relative smaller deformation ability was observed by comparing with arc heating under same power scale.
4. The above three different heating processes were all successfully simulated with the developed heat sources.

References

1. Park JU, Park SC, Lee CH (2009) Control of longitudinal bending distortion of built-up beams by high-frequency induction heating. Weld J 88:29–34
2. Radaj D (2012) Heat effects of welding: temperature field, residual stress, distortion. Springer Science & Business Media, Berlin
3. Tsirkas SA, Papanikos P, Kermanidis Th (2003) Numerical simulation of the laser welding process in butt-joint specimens. J Mater Process Technol 134(1):59–69
4. Moraitis GA, Labeas GN (2009) Prediction of residual stresses and distortions due to laser beam welding of butt joints in pressure vessels. Int J Press Vessels Pip 86(2–3):133–142
5. Wang J, Yuan H, Ma N, Murakawa H (2016) Recent research on welding distortion prediction in thin plate fabrication by means of elastic FE computation. Mar Struct 47:42–59
6. Goldak J, Chakravarti A, Bibby M (1984) A new finite element model for welding heat sources. Metall Trans B 15(2):299–305
7. Azar AS, Ås SK, Akselsen OM (2012) Determination of welding heat source parameters from actual bead shape. Comput Mater Sci 54:176–182
8. Xia H, Li L, Ma N, Huang H (2016) Analysis of influence of initial gap of laser-arc hybrid butt welds on inherent strain and welding deformation of a pipe structure. J Laser Appl 28 (2):022009

The Comparison of Different Models of Government Purchasing Home-Based Aged-Care Services

Guoping Zhang

Abstract By taking typical examples from three different districts (Haishu in Ningbo, Gulou in Nanjing and Canglang in Suzhou), the article analyzes and discusses the differences in terms of the policies and mechanisms of the local governments' purchasing home-based aged-care service in order to reveal different mechanisms, efficiency features and development levels in these districts.

Keywords Government · Purchasing home-based aged-care service
Policy · Operating mechanism · Haishu district in Ningbo · Gulou district in Nanjing · Canglang district in Suzhou

1 Introduction

The government purchasing public services is a mechanism by which the government provides funds and the social organizations take the contracts so that specific public services goals are realized based on the contractual relations. In 2001 the Ministry of Civil Affairs initiated nation-wide the community welfare services for the aged, the Starlight Program. Since then, all communities and districts have been exploring and implementing government purchasing programs and policies considering their various situations with the local support to develop home-based aged-care service programs such as providing free services and subsidies for those old-aged who are poor, with special contributions or the oldest old, and establishing policies of financial funds purchasing services, service organizations providing services and home-based old-aged residents accessing services, and thus formed new mechanisms of government purchasing home-based aged-care services. Haishu in Ningbo, Gulou in Nanjing and Canglang in Suzhou are three typical and pivotal local districts that adopt government purchasing home-based aged-care services. The article compares the core contents of the models adopted by the above three

G. Zhang (✉)
School of Management, Changshu Institute of Technology, Changshu, China
e-mail: gpz@cslg.edu.cn

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_33

districts from the perspectives of levels of competition, inter-subjective relationship, levels of institutionalization and quantities and scales of purchases.

2 The Level of Competition of the Government Purchasing Home-Based Aged-Care Services

The level of competition, a term to measure marketization of the government purchasing public services can be divided as competitive and non-competitive, and can be inspected through the number of organizations participating in the competition and the methods that they participate in.

When considered in terms of the government purchase, the above three examples are all insufficient as they adopted government-entrusted operation. The reason for that is there are few social organizations capable of undertaking the tasks and little investment from the government sources, and thus the related government agencies such as offices of Civil Affairs and offices of Committee on Aging pick the operation units based on information available. Taking Haishu in Ningbo for example, the Starlight Home for the Aged Association in Haishu is a social organization established in June of 2003 with the purpose of undertaking home-care for the aged services. It is different from the ordinary associations of the same kind as it is supported by the district committee and the government. The Aged Home service center in Gulou district was established in 2001 based on a well-known private nursing home in Nanjing, and has been focusing on the old-aged people's needs to develop a series of influential activities favored by the local old people. These form the basis of mutual trust between governments and non-governmental organizations, thus making Gulou's service center the best suitable contractor. Suzhou Dingsheng Property Management Corp. Ltd, the major operator of Canglang Virtual Nursing Home, is a professional and modern property management enterprise with management qualification of level two. Dingsheng was originally a state-owned enterprise with close ties with the government.

The entrusted purchase is oriented and non-competitive. The selection procedure without competition leaves room for development for the assigned non-governmental organizations, but the monopoly of resources will prevent the rest outstanding organizations from entering the field, participating and providing public services. Also, since these assigned non-governmental organizations do not go through severe competition, they are likely to become sluggish in improving their work for lack of competitors and thus causing deterioration of service quality and low efficiency.

Philosophy and Social Science Planning Projects Organized by Jiangsu Province (15SHA001); Philosophy and Social Science Planning Projects in Universities Organized by Jiangsu Province (2016ZDIXM044).

3 Inter-subjective Relationship

The inter-subjective relationship, or the relationship between the contractor (non-governmental organizations) and the contractee (governments), can be divided as independent and subject. Considering the three purchasing models of Haishu, Gulou and Canglang, the cooperative relationship between the governments and the social organizations can be defined as an independent one because the contractor (non-governmental organizations) is a lawfully established, independently operated and developed, and self-disciplined incorporated organization and working unit. However, since differences exist in terms of organizations' initiators, important decision-makings, appointments and sources of funds, some different aspects in such relationship's independence are demonstrated in the purchasing models of Haishu in Ningbo, Gulou in Nanjing and Canglang in Suzhou.

Considering Haishu model, Starlight Home for the Aged Association is an independent social organization which has been cultivated by the government and is purely for the purpose of the completion of the governmental functions as well as the public administrative affairs. The government provides offices, necessary equipment and a three-hundred-thousand budget annually. Cui Dehai, the retired Minister of Propaganda Office in Haishu, serves as its president. Thus, it is a social organization with strong governmental background and close ties with the government. It depends on the government financially which means it cannot be without influences or even interventions from the government, not much though, in its daily operation, and it then cannot be truly organized and administrated independently and healthily. The service center's development and administration are restricted by and dependent on the government, which may lead to its becoming affiliated with the government as a branch. This will, rather than relieving the government's burden as originally expected by the government, expand potentially the governmental functions, and finally cause the abuse of the public resources. The lack of independence and decision-making capability will also lead to the slow development of the service center's self-administration and self-development, or even degeneration. If this remained long, the quality of the service would gradually be degraded, which will affect the long term promotion and implementation of the home-based aged-care services policy.

The major operators in Canglang's model are the Neighborhood Emotion Virtual Nursing Home as well as the commercial entity, Suzhou Dingsheng Property Management Corp. Ltd. The nursing home, as a social organization, is a non-governmental non-profit organization with governmental and commercial backgrounds. The operating entity is actually under the administration of both the non-governmental non-profit organization (the virtual nursing home) and the commercial enterprise (the property management corporation) whose cooperation formed a philosophy of market-oriented operation, enterprise management and professionalized service. The independence of its operation can be clearly attributed to the increasing number of self-funded old-aged people and the expanding services. For instance, the virtual nursing home, apart from guaranteeing free services

to those supported by the government, extends the target and scope of its service to self-paying clients to ensure enough funding. It also varies the service items to include the residents, absorb some food and oil companies of certain scale and credibility, and establish the logistics center to provide delivery services to ordinary residents.

The Aged Home Service Center in Gulou district of Nanjing is similar to the Virtual Nursing Home in its operation mechanism to some extent, but not entirely the same. The similarity lies in that both of the two entities are supported by the governments, and in addition, they undertake the contracts from the governments based on their previous businesses, expand their services and have established the non-governmental non-profit organizations. Since the new entities originate from the old ones, the inherited features and resources such as the philosophy of operation, operational mechanism, and management experience provide the support and basis in terms of human resources, financial capital and supplies for the operation and development of such non-governmental non-profit enterprises. Gulou's model differs from Canglang's in that it, a combination of non-governmental non-profit enterprises, i.e. the Aged Home Service Center and the non-governmental aged-care organizations, is based on the non-governmental organizations, which are also public welfare organizations, so it does not have a clear planning on commercialization and operation as compared with the virtual nursing home.

4 Levels of Institutionalization of the Government Purchasing Homed-Based Aged-Care Services

The level of institutionalization can be used to measure the stability and trend of the current government purchasing, which can be judged by inspecting the laws and documents, the procedures and the abundance of the supplies of the resources. The government purchases of public services in our country currently lack the support of laws on the level of the nation. The government purchases of the home-based aged-care services were all initiated by the local governments, the lawful basis of which are mainly from local policies and regulations. These documents are not enough quantitatively and not authoritative, as they are issued and implemented by the local governments of the district level.

For Haishu in Ningbo, the government purchase is based on the following policies and regulations. (1) In December of 2004, local government in Haishu district released Guidelines on the Work of Home-based Aged-care in Haishu District which suggested that the local government address the difficulties of those old-aged who cannot take care of themselves, or who are not taken care of by their relatives and thus need daily care services by government purchasing services. (2) The Proposal of Promoting the Building of Aged Care Services System, issued by Ningbo in August of 2010 suggested that it be to strengthen the leadership and the investment of the system construction, promote the construction of the service

points, increase the support of policies for non-governmental institutions and reinforce the team building of the aged-care service. (3) The Opinions on Further Strengthening Implementation of Home-based Aged-care Services, issued by Haishu district in July of 2011 required proper enlargement of the services provided by the government and improvement of the government purchase mechanism.

For Canglang in Suzhou, the policies and regulations include: (1) Opinions on Suzhou's Speeding Up the Development of Aged Care Services, issued in December of 2005, required that governmental financial departments support aged-care institutions and home-based aged-care services organizations financially, government implement assistance services for those who are at home but with difficulties, and set up mechanisms of investment in aged care cause. (2) Opinions on Speeding Up Aged Care Program in Canglang District was formulated in 2007, and On Promoting the Implementation Program of Neighborhood Emotion Virtual Nursing Home, and Measures on Subsidizing Aged Care Service Organizations in Canglang District formulated both in 2008. Canglang district brought the virtual nursing into the aged-care service organizations so the virtual nursing can be entitled to subsidies of setup and operation provided by the government. The virtual nursing home obtains funds from government, profits from services and collections from societies. For those qualified for governmental assistance, different levels of purchasing services are implemented as 450 Yuan, 350 Yuan, 250 Yuan, 60 Yuan and 30 Yuan per person per month. 3. In February of 2010 Suzhou formulated Supplemented Opinions on Further Speeding Up Aged Care Services Cause in Our City, which increased subsidies of home-based aged-care services and numbers of those qualified, adjusted the standards and items of assistance, strengthened the investment mechanisms and completed regimes of various aged-care services.

The government purchasing services of Gulou district in Nanjing are based on these policies and regulations: (1) Gulou district formulated in 2005 The Implementation Program of Gulou Home-based Aged-care Services, followed by The Implementation Program of Scale Enlargement of Home-based Aged-care Services Network, which require further increasing public financial investment for the construction of the network. (2) For further improving mechanisms of evaluation and supervision, the government formulated Evaluation and Supervision Programs of Home-based Aged-care Service Network to regulate in details the teamwork building, operation process, feedback, training and management, and meantime to form daily supervision teams by the Old-aged Community Association. (3) Approaches to Financing Non-governmental Organizations and Individuals to Establish Social Welfare Institutions in Gulou District was issued in 2006 with the purpose of supporting and subsidizing newly established social welfare institutions and community aged-care centers.

In comparison of the above three districts, Suzhou and Nanjing made more detailed the relevant legal documents concerning social organizations involved in the home-based aged-care services. As judged from the level of governmental purchases, it is higher in Suzhou than in Nanjing and Ningbo. The government in Suzhou subsidizes annually 20–100 thousand Yuan to day-time nursing centers, aged-care centers. The government in Nanjing provides 10 thousand Yuan to each

aged-care center and service center, subsidizing 60 Yuan to each old-aged person monthly. The government in Ningbo offers at least 20 thousand Yuan to each urban community home-based aged-care service center that has functions such as living services, entertainment and education, and at least 10 thousand Yuan to each rural community Starlight Home for the Aged. The government in Suzhou grants every month a 450-Yuan coupon to those qualified for governmental assistance, and among these who are qualified the taken-cared persons will receive coupons each worth 600 Yuan. The government in Nanjing provides free homemaking services amounting to 20 h each month to qualified aged residents (those of venerable age, of living alone and of difficulties including physically disabled). The government in Ningbo provides 1-h free homemaking services each day to those qualified old-aged persons living alone without children.

5 Quantities and Qualities of the Services

Since it is not long before the governments started purchasing home-based aged-care services, the relevant policies made by the governments of all levels are only exploratory, which is shown in the insufficient funding, small numbers of people entitled and fewer items of services provided. By 2011, the old-aged persons who are covered by the services in Haishu are counted about 600 or so, less than one percent of all the old-aged people. Even if those who are assisted by volunteers and who are self-paid are included, the number will only rise to about four percent. Gulou district has increased the target of the aged-care service network to include some empty nesters and some old-aged with difficulties from previously those living alone. By 2009, 2500 aged residents in Gulou district had received free services, paid- and discounted services, which have covered 1600 households and more [1]. The total number of those serviced is over 5000, which accounts for about 5.3% of the population of the aged over 60 years old. Canglang's Neighborhood Emotion Virtual Nursing Home had been covering 456 households by October, 2007, and by February, 2011 the number rose to 3620 (907 households, 26% of the total, are paid clients) households, among which nearly 6000 empty nesters, 10.3% of the total aged people were included.

The services mainly include life care, recovery nursing and spiritual comforting, but the offerings are limited to simple caring and home-making services, lacking medical services and spiritual comforting, which, even if included in the content of the services, are not high quality or satisfactory.

By comparing the three models and their operation mechanisms, it is found that the mechanisms of the non-governmental organizations' operations can be divided into three models, which are namely government-social organization-community as called Ningbo Haishu Model, government-social organization-private nursing home-community service center as called Nanjing Gulou Model, and government-social organization-commercial enterprise-community service center as called Suzhou Canglang model. The three models differ in the following way.

Haishu model's operation entity, Starlight Home for the Aged Association, is attached to the government, Gulou model's, Aged Home Service Center, attached to private owned aged-care nursing homes, and Canglang model's, Virtual Nursing Home, attached to property management enterprises. The entities to which they are attached determine the operational model and institutional efficiency.

Although the government purchasing home-based aged-care services made great achievements in experimental and exploratory areas, the overall effect is limited. From the three cases in the article, the local governmental purchases are only supplemental and relieving, benefiting those venerable aged and living alone who only take up a small portion of the total. For those who are venerable aged, weak and sick, and not capable of taking care of themselves, home-based aged-care does not meet the requirement, for their living can't sustain without full-developed and continuous daily life care services, i.e. the highly needed community long-term caring and nursing system, especially focusing on disabled and semi-disabled, is extremely deficient. The current movement of such purchase reform is inspired by some local governments, rather than an overall formal institutional arrangement of home-based aged-care services. From this perspective, the existing feasibility will obviously remain limited in the foreseeable future, and thus institutional construction should be the key content of government purchasing home-based aged-care services.

Reference

1. FAN W-F et al (2010) Government purchases of public services from CSOs on the home-based aged-care services: a case of Gulou district in Nanjing. Scientific Decision

Study on Value Increment of Channel Value Chain Between Manufacturer and Distributor-Based on Game Theory

Junwei Ma and Jianhua Wang

Abstract In the fierce market competition, how to carry out strategic coordination effectively and achieve the increment of value, which has become an important basis for manufacturers to maintain a strong competitive vitality. This empirical analysis shows that it can help manufacturers realize the appreciation of channel value chain and improve the whole industrial operating efficiency so as to promote the ability of both manufacturers and distributors by setting up fair profit distribution point, strengthening communication, playing the role of penalty and market reputation mechanism, and establishing warning mechanism.

Keywords Channel value chain · The relationship between manufacturers and distributors · Game theory

1 Introduction

Manufacturers and distributors not only compete but cooperate with each other, which is a contradictory benefit community. In the fierce market competition, how to make strategic synergy efficiently and realize the value increment of channel chain becomes the important foundation for manufacturers to maintain strong market competition vitality.

Scholars regarded the structure, behaviour and performance of channel as the focus of marketing channel research, and introduced the channel system and its social inevitability from two angles of institution and function of channel. Western scholars are more willing to call the channel network as the system, and believe that all functions of the channel system can be implemented within the system. Therefore, the focus of channel research has gradually shifted to the behaviour and relationship among members of the channel system.

J. Ma (✉) · J. Wang
Changshu Institute of Technology, No. 99, 3rd South Ring Road, Changshu, China
e-mail: horse0628@163.com

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_34

In the current researches, scholars have begun to use value chain to study channel efficiency, and have paid close attention to the externalization and marketization of enterprise internal value chain. At the same time, based on channel relationship theory, scholars gradually explored the factors affecting relationship between manufacturers, and proved the effect of these factors through scientific testing and other methods, which provided a foundation for channel relationship management.

From the point of view of cooperation, through the analysis of the economic behaviour and transaction characteristics between manufacturers and distributors, as well as assumption of model conditions, this study established a mixed game model based on mutual cooperation between manufacturers and distributors in order to explore the factors which affect the cooperation. At the same time, this study constructed a finite repeated game model so as to seek the conditions for long-term cooperation.

2 Literature Review

In regard to channel structure, with the development of social economy and the drastic change of market environment, scholars proposed the three-dimensional and flat channel system. The scholars focused on three aspects of point (critical point chose in city marketing channel system), line (the route of commodity circulation), surface (radiation range formed by lines and points) in order to promote the marketing key power to form an efficient channel system with reasonable association and aggregation in the distribution. At the same time, the scholars modified and improved the traditional vertical channel structure so as to solve problems such as control difficulty, information transmission distortion, low efficiency, overstaffed organization and other issues and deal with channel conflict. In order to improve the efficiency of the whole channel system, scholars have believed that the channel system should be as flat and flexible as possible.

In terms of channel members' behaviour, scholars focused on the power, conflict and cooperation among channel members. Meanwhile, scholars believed that power, conflict and cooperation arose from the industrial dependence between organizations. The revenue of channel members depends on whether they can sell the product successfully to the target customer or not. Therefore, manufacturer's power is related to the product it controls and distributor's power is related to the channel he controls. According to the sources of power, scholars divided channel power into six kinds: reward, legal, identity, compulsion, expert and information. In order to facilitate the study, some scholars had divided the power of manufacturers into two categories, namely, coercive power and non coercive power.

About the relationship between channel members, scholars believed that although the cooperation between manufactures and distributors often terminated due to conflicts of interest, but it was helpful to enhance communication and negotiation between manufactures and distributors as well as can help solve the

problem rapidly, if the stable cooperation maintained in the channel system. After 1990s, marketing management experts put forward the theory of relationship marketing, and studied the performance, purpose, essence and life cycle of the whole channel relationship. Scholars had conducted a more detailed study of the establishment, maintenance and application of the channel relationship. From the perspective of the relationship between retailers and manufacturers, Madeline and Betsy (1999) pointed factors that influenced the cooperation such as communication, power, internal environment and leadership style. They also proposed some effective measures to maintain the relationship between channel members [1]. According to the level of cooperation and binding force of contract, Stern and Ansary (1992) divided channel system into conventional type, management type and contract type, which clarified the internal relations existed in different channels as well as coordination and management in channel system [2]. According to the uncertainty of market environment, additional value in industrial downstream and manufacturers' substitutability, Keysuk and Gary (1996) divided marketing channel system into eight situations, and pointed out the joint strength of each kind of situation which manufacturers and distributors need in order to maintain cooperation [3]. Zhuang (2002) did the empirical analysis on channel behaviour of Industrial and Commercial Bank of China based on the channel behavior theory, which discussed the relationship between the use of coercive power, non coercive power, conflict and cooperation [4]. Scholars such as Zhuang et al. (2008) also examined the relationship between channel governance and interpersonal relationship among organizations, and came to a conclusion that the level of interdependence between channel members, non symmetry and interpersonal emotional relations across the organization will affect the speculative behaviour [5]. From the perspective of game between manufacturers and distributors, Liao and Liu (2003) pointed out that good cooperation of manufacturers and distributors in channel system is an effective means for them to advantages complementation, adaptation to the changing market environment and realization the value increment of channels. During the cooperation, it had to increase trust and strengthen mutual binding between manufacturers and distributors in order to strive to maintain a win-win situation [6]. Through the establishment of game model, Xiao (2008) studied the relationship between manufacturers and distributors in Chinese IT storage market, and pointed out that it will help to establish benign cooperative game system with the construction of incentive system, improvement in the symmetry of market information and establishment of integrity mechanism [7].

3 Case Analysis and Propositions

In this study, it proposes a mixed game model based on mutual cooperation between manufacturer and distributor in order to explore the factors which affect the cooperation. After that, it gives some propositions in order to keep better relationship and achieve win-win situation.

3.1 Model Assumptions and Definitions

This study used the game theory to analyze value chain of channel. The participants mainly were manufacturers and distributors. In order to make it easier to analyze and define the relevant concepts in the model more accurately, the following assumptions and definitions were made:

Hypothesis 1: suppose that in order to build distribution channels, manufacturer 1 considers cooperating with distributor 2 and the number of cooperation between the two parties is not limited. Of course, they can either cooperate with each other or not (the first choice of non cooperation between the two parties is not betrayal), and then manufacturer 1 and distributor 2 are both participants in the game model. If manufacturer 1 considers building channels by itself, the channel departments will be regard as distributor 2 as well.

Hypothesis 2: suppose that manufacturer 1 and distributor 2 reach consensus on cooperation, they negotiate with each other and sign contract. If both parties are responsible for the completion of product transaction in accordance with the contract, the two sides will cooperate successfully. At this point, manufacturer 1 accesses to direct economic benefit R_1^c and distributor 2 obtains direct economic benefits R_2^c, S_1 and S_2 are the gains for each party due to the promotion of their own ability during the cooperation. For example, because of cooperation, manufacturer' manufacturing capacity and innovative ability will be improved, and distributor' market development and service capabilities will also be improved to some extent.

Hypothesis 3: if one of the parties breaks the contract, it will be punished and the fine is P. Because the default effects on manufacturer and distributor are different, so the size of their damage can be defined with the fine multiplying an influence coefficient. Consider making a definition: if manufacturer 1 breaks contract, the amount of its penalty is k_1P, and in the same way if distributor 2 breaks contract, the amount of its penalty is k_2P. Take into consideration that both of them may have betrayal together, defaulter' non-physical loss such as the reputation will increase, which will affect the future income of the defaulting party. Therefore, non-physical loss caused by betrayal is as follows: the intangible loss for manufacturer is C_1, and the intangible loss for distributor is C_2.

Hypothesis 4: manufacturer 1 and distributor 2 are semi-rational participants. Therefore, if there is no external interest to induce them, we can not take into account the default situation. If a party defaults, the defaulting party will obtain additional income. It is assumed that additional income for the manufacturer 1 is R_1^d, additional income for the distributor 2 is R_2^d.

Hypothesis 5: if one of the participants has a default, the other side will cause loss (mainly asset exclusion loss and the cost of the preparation before the cooperation). If there is a default from manufacturer 1, the loss for distributor 2 is R_2^{ch}(Because loss R_2^{ch} is a positive real number, it is hereby announced that equivalent symbols L_2^{ch} is defined. It is marked as benefit when more than 0; otherwise it is marked as loss. The same applied below). If distributor 2 breaks the contract, the loss for the manufacturer 1 is R_1^{ch}. If betrayals happen to both parties at

the same time, the loss of each party shall be the same as that each has only when one betrayal occurs.

Hypothesis 6: assume that both parties are involved in signing the contract, but it is unknown that whether the two parties fulfil contract or not in accordance with the contract stipulations. It is also assumed that the probability of contract execution by the manufacturer 1 is θ, that is to say, the probability of contract breach is $(1-\theta)$; and the probability of contract execution by the distributor 2 is β, the probability of the contract breach by the distributor 2 is $(1-\beta)$.

3.2 Model Analysis and Recommendations

According to the hypotheses, the game relation between manufacturer and distributor in the channel is a mixed strategy Nash equilibrium. Then, manufacturer 1's hybrid strategy is $\sigma_1 = (\theta, 1-\theta)$, and distributor 2's hybrid strategy is $\sigma_2 = (\beta, 1-\beta)$. The profit model for the two participants in cooperative game is shown in Table 1:

It shows that the expected profit function of the manufacturer is:

$$V_1(\sigma_1,\sigma_2) = \theta\left[\beta\left(R_1^c + S_1\right) - (1-\beta)L_1^{ch}\right] + (1 - \theta)\left[\beta\left(R_1^d - k_1P - C_1\right) + (1-\beta)\left(R_1^d - k_1P - C_1\right)\right] \quad (1)$$

That is:

$$\frac{\partial V_1}{\partial \theta} = \beta\left(R_1^c + S_1\right) - (1-\beta)L_1^{ch} - \beta\left(R_1^d - k_1P - C_1\right) - (1-\beta)\left(R_1^d - k_1P - C_1\right) - \beta\left(R_1^c + S_1 + L_1^{ch}\right) - \left(L_1^{ch} + R_1^d - k_1P - C_1\right) \quad (2)$$

Set $\frac{\partial V_1}{\partial \theta} = 0$, we can get:

$$\beta^* = \frac{\mathrm{R_1^d} - \mathrm{k_1P} - \mathrm{C_1} + \mathrm{L_1^{ch}}}{R_1^C + S_1 + L_1^{ch}} \quad (3)$$

Table 1 Profit model of cooperation game between manufacturer and distributor

Manufacturer's strategy	Distributor's strategy	
	Cooperation β	Betrayal (1 − β)
Cooperation θ	$R_1^c + S_1$, $R_2^c + S_2$	$-L_1^{ch}$, $R_2^d - k_2P - C_2$
Betrayal (1 − θ)	$R_1^d - k_1P - C_1$, $-L_2^{ch}$	$R_1^d - k_1P - C_1, R_2^d - k_2P - C_2$

Similarly, the expected profit function of the distributor is:

$$V_2(\sigma_1,\sigma_2) = \beta\left[\theta\left(R_2^c + S_2\right) - (1-\theta)L_2^{ch}\right] + (1 - \beta)\left[\theta\left(R_2^d - k_2P - C_2\right) + (1-\theta)\left(R_2^d - k_2P - C_2\right)\right] \quad (4)$$

set $\frac{\partial V_2}{\partial \beta} = 0$, we can get:

$$\theta^* = \frac{\mathrm{R_2^d} - \mathrm{k_2P} - \mathrm{C_2} + \mathrm{L_2^{ch}}}{R_2^C + S_2 + L_2^{ch}} \quad (5)$$

Therefore, the mixed strategy Nash equilibrium in the game for manufacturer and distributor is: $\sigma^* = \sigma(\theta^*, \beta^*)$.

In addition, if the probability of choosing the cooperation strategy for distributor is β, then the probability of choosing betrayal strategy is $(1-\beta)$, the expected return of cooperation strategy for manufacturer is:

$$V_1(1,\beta) = \beta\left(R_1^C + S_1\right) - (1-\beta)L_1^{ch} \quad (6)$$

And the expected return of betrayal strategy for manufacturer is:

$$V_1(0,\beta) = \beta\left(R_1^d - k_1P - C_1\right) + (1-\beta)\left(R_1^d - k_1P - C_1\right) \quad (7)$$

If we want the manufacturer to choose cooperation strategy, it must ask for

$$V_1(1,\beta) \geq V_1(1,\beta) \quad (8)$$

By putting (6) and (7) into (8), we can get:

$$\beta \geq \frac{R_1^d - k_1P - C_1 + L_1^{ch}}{R_1^C + S_1 + L_1^{ch}} \quad (9)$$

It means that if $\beta \geq \beta^*$, the choice of manufacturer's strategy is cooperation. On the contrary, the choice of strategy is betrayal.

If the probability of choosing the cooperation strategy for manufacturer is θ and the probability of choosing betrayal strategy is $(1-\theta)$, then the expected return of cooperation strategy for distributor is:

$$V_2(\theta,1) = \theta\left(R_2^C + S_2\right) - (1-\theta)L_2^{ch} \quad (10)$$

The expected return of betrayal strategy for distributor is:

$$V_2(\theta,0) = \theta\left(R_2^d - k_2P - C_2\right) + (1-\theta)\left(R_2^d - k_2P - C_2\right) \quad (11)$$

If we want the distributor to choose cooperation strategy, it must ask for:

$$V_2(\theta, 1) > V_2(\theta, 0) \tag{12}$$

That is to say, $\theta > \theta^*$

$$\frac{\partial \beta^*}{\partial L_1^{ch}} = \frac{\left(R_1^c + S_1\right) - \left(R_1^d - k_1 P - C_1\right)}{\left(R_1^c + S_1 + L_1^{ch}\right)^2} \tag{13}$$

$$\frac{\partial \theta^*}{\partial L_2^{ch}} = \frac{\left(R_2^c + S_2\right) - \left(R_2^d - k_2 P - C_2\right)}{\left(R_2^c + S_2 + L_2^{ch}\right)^2} \tag{14}$$

It can be seen from Eqs. (13) and (14) that the expected return of the cooperation between the two parties is greater than that of the choice of both betrayals, if $\left(R_i^c + S_i\right) - \left(R_i^d - k_i P - C_i\right) > 0$. At the moment,

$$\frac{\partial \beta^*}{\partial L_1^{ch}} > 0, \frac{\partial \theta^*}{\partial L_2^{ch}} > 0$$

As we can see, when the R_i^c and S_i increasing and R_i^d, $k_i P$ and C_i decreasing, $\theta*$ and β^* become bigger and bigger. It needs a stronger willingness to cooperate with each other. If the manufacturer 1 and distributor 2 complete the contract and reach cooperation, the value range of θ and β should be greater (that is, the value is closer to 1).

At this time, if the temptation outside of the system becomes bigger, and the cooperation income as well as the default penalty becomes smaller, and the invisible loss caused by a breach of contract becomes smaller, both sides need to have greater confidence to promote cooperation. For example, the manufacturer's competitors may induce distributor to terminate its cooperation with the original manufacturer by a higher share or benefit. At this point, once the distributor defaults, the distributor will be punished with small fines compared to its benefits and if the loss of its reputation is not very big, it can easily lead to the breakdown of previous cooperation. Therefore, it needs to show more sincerity and let both sides can recognize each other's sincerity during the cooperation in order to deepen and strengthen cooperative partnerships. When the cooperation income is greater than the betrayal earnings, direct economic benefit due to cooperation and the benefit because of enhancement of their professional ability are greater than that of betrayal, pure strategy Nash Equilibrium of participants in cooperative game is (cooperation, cooperation). At the same time, from the result of Table 1, we can see that both sides will choose cooperation as their dominant strategy. However, if one side breaks the contract and causes the other party to suffer a great loss, then the side suffered with larger loss will choose the partner with higher confidence coefficient before the cooperation in order to complete the transaction

In real life, cooperation between manufacturer and distributor is more represented as a long lasting cooperative relationship. In general, the pursuit of profit maximization is the ultimate goal of their long-term cooperation for both manufacturer and distributors. Therefore, regardless of how much attention they pay to the benefit of a separate collaboration with each other, they consider the total benefits of long-term cooperation more important. Of course, due to changes in economic environment and market instability, as well as internal business changes in manufacturer and distributor, continual and indefinite cooperation between them is really hard. The fact is that many of the two sides only carried out limited cooperation. Therefore, the long-term cooperation game between manufacturer and distributor can be viewed as a finite repeated game process.

Based on the principle of maximizing their profits, manufacturer and distributor can choose strategies for cooperation or non cooperation. However, in the course of repeated cooperative games, the cooperation between manufacturer and the distributor will cease as soon as one side chooses not to cooperate or has a non cooperative action. At this point, the manufacturer 1 will establish a new distribution channel and distributor 2 will choose other manufacturer or carry out other projects.

In order to facilitate the analysis, it is assumed that the manufacturer chose the non cooperative strategy at the beginning, and the results of the manufacturer and distributor were (R_1, R_2). If both sides choose the cooperation strategy and fulfill the contract in accordance with the regulations, the first cooperation between the two sides is successful. Then, the two sides continue to cooperate with each other, which will result in the proceeds of $\left(R_1^{ci}, R_2^{ci}\right)$. If one or both sides fail to comply with the contract and assume a breach of contract, the cooperation between the two parties shall fail or terminate. Due to the choice of the betrayal strategy, the expected return of the defaulting party is: $R_j^{di} - k_j P - C_j$, therein, $i = 1 \ldots n, j = 1, 2$.

The party who chooses the cooperation strategy will make a loss because of the default of the other party. Its benefit will be: $-L_m^{ch}$, therein,$m = 3 - j$.

The cooperative game relation model can be expressed in Fig. 1 (finite repeated game).

We can see from Fig. 1:

1. If $R_j^{ci} + S_j^i > R_j^{di} - k_i P - C_j^i$ then, the best strategy for both sides is to cooperate and the possibility of establishing a long-term cooperative relationship between the two parties will exist.
2. If $R_j^{di} - k_i P - C_j^i > R_j^{ci} + S_j^i$, the two sides will choose the strategy of non cooperation, the cooperation between the two sides can not continue.
3. If $R_j^{ci} + S_j^i << R_j^{di} - k_i P - C_j^i$, the choice of betrayal is the best strategy of the game, the traitor gains greater benefit because of short-term opportunism.
4. If $R_j^{ci} + S_j^i < R_j^{di} - k_i P - C_j^i$, the difference is small, the participant J' strategy may be varied. For example, when the manufacturer has established a good partnership with distributor, although the participant J' income is not high or

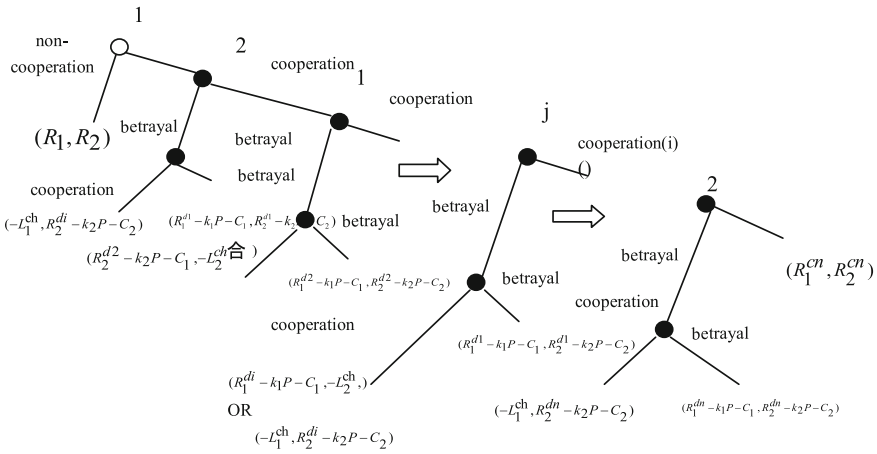


Fig. 1 Repeated game profit model between manufacturer and distributor

even a loss in the i(th) time cooperation, but this is only temporary compared to the long-term relationship between them. Therefore, the two sides will continue to cooperate in the long run, mainly considering the expected profits of long-term cooperation, which can compensate for the loss. On the contrary, they will choose betrayal [6].

From the repeated game analysis between manufacturers and distributors, we can see that the long-term cooperation between manufacturers and distributors is based on the pursuit of more benefits. Therefore, whether we can maintain the long-term interests of the channel members is a prerequisite for cooperation. When bilateral cooperation is unprofitable, the long-term cooperation of manufacturer and distributor will be lacking in foundation. In addition, in the fierce competitive environment, even if cooperation is high frequency with stable relationship, but as long as a breach of contract or betrayal, the two sides may not continue to cooperate with each other. Therefore, in order to ensure stable long-term cooperation among channel members, it is necessary to strengthen channel incentives and cooperation trust.

Conclusion 1: the "win-win" situation and the promotion of channel system integration ability are the preconditions for the value-added of channel value chain.

Game results show that "win-win" is a prerequisite for cooperation. Through the cooperation, manufacturer and distributor can share the profits and the income is greater than other investment or speculative gains, on the basis of which the cooperation between the two sides with long-term stability can be realized. Therefore, when both sides make full use of their own resources and create revenue, they should also establish a platform for cooperation and actively utilize the system resources generated by the cooperation between the two sides. Only the cooperation between the two sides and the value creation in the system resources could have

improved the integration ability of channels and realized the value increment of channel chain.

Conclusion 2: strengthening understanding and mutual trust as well as promoting the feedback ability of information in channel system is the basis of value increment of channel chain.

Game results show that strengthening understanding and trust is the basis of improving the relations between manufacturer and distributor. In the channel value chain, the breakdown of cooperative relationship between manufacturer and distributor is related to communication mechanism, trust mechanism and mutual benefit mechanism to a great extent, which leads to the lack of the basis for the establishment of long-term cooperative relationship. Therefore, understanding, mutual trust and good communication among channel members are conductive to the improvement of cooperative relations so as to enhance the information feedback ability of the channel system and realize the value-added of the channel value chain.

Conclusion 3: establishing market reputation mechanism, improving the penalty for breach of contract, and enhancing the proprietary ability of manufacturer and distributor are the powerful guarantee to achieve value-added of channel value chain.

The results of the game model show that although the choice of betrayal is profitable in the short term, the intangible loss is greater. If the channel members know or predict a greater loss of punishment due to breach of contract, the possibility of cooperation will be greater. The pursuit of self-interest and disruption of channel cooperation are often the important reasons leading to the end of strategic cooperation among manufacturer and distributor. Therefore, only through the reputation mechanism and increment of the penalties, so that it is not worth to break the contract for the defaulters, which can effectively help establish lasting cooperation relations. Larger restraint and strong reliability of bilateral cooperation are useful to enhance their proprietary ability and achieve the value increment by making high default penalty in the contract as well as strengthening supervision and completing industrial association statute.

Conclusion 4: resolving the risk of cooperative operation and enhancing the strategic synergy ability are the key point to realize the value-added of channel value chain.

The result of the game shows that reducing or sharing the risk of cooperation in the process of cooperation is the key to the cooperation between the two parties. In view of exclusive resources from both sides, high investment of manufacturer or distributor in the process of cooperation often contains high risk. Whether the risk is able to effectively be reduced or shared is the key problem, which usually influences further cooperation. Therefore, manufacturer and distributor can reduce the operational risk through the strategic coordination of capital integration, club model and contract model, so as to realize the value-added of the value chain.

4 Conclusions

Through constructing the game model of channel value chain, this study analyzed the conditions and main influencing factors of value-added for channel value chain, and also established a repeated game model of multiple cooperation. Through the analysis, it pointed out that "win-win" cooperation, confidence consolidation and punishment enforcement as well as cooperative risk relief are the effective means to solve the inefficient operation and strong guarantee to promote the value increment of channel value chain.

These conclusions have important practical significance. In order to realize the value-added of channel value chain, the channel system should focus on the profit distribution, cooperation power, binding force and cooperation risk of each other.

First, we should set up a fair profit distribution point. All enterprises in the channel system should explore how to set a fair profit or reward distribution point through consultation in order to encourage them to do their own professional work, realize the efficient operation of the whole system and win the competitive advantage in the industry. Only when the profit distribution is more reasonable, adjustable and more attractive, cooperation can be achieved with win-win situation.

Second, mutual understanding and communication must be strengthened. Reasonable personnel management and communication mechanism should be established with electronic information system in order to promote information exchange, which will help reduce strategic misjudgment on both sides and eliminate tension and distrust in order to enhance information feedback and channels coordination.

Third, it should play the role of high penalty for breach of contract and the reputation mechanism. High default payment can reduce the enterprise's betrayal tendency. Strengthening supervision, completing the industrial association statute, and playing the role of reputation mechanism can not only improve the ability of self-control, but also can protect proprietary resources are fully utilized in each link of the channel system, which can let both manufacturer and distributor pay attention to the development of their professional ability and industrial competition.

Forth, it is very important to establish early warning mechanism. Strengthening information processing capacity and establishing warning mechanism are powerful guarantee for the timely and accurate understanding of the operational risks by the channel system. Through early warning system, the channel system not only can get enough time for it to resolve the crisis and deploy enterprise resources with joint efforts. Only by establishing an early-warning mechanism among the manufacturer and distributor, can they jointly guard against the risk of cooperative operation in the channel system and realize the value-added of the channel value chain.

Acknowledgements This research was financially sponsored by fund of six talents peak in Jiangsu province (Grant No. JY-001),Qing Lan Project of Jiangsu Province in China (Grant No. SZ2014003), Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province (Grant No. 2016SJB790045).We would like to thank anonymous references

for their insightful comments and suggestions which lead to the significant improvement and better presentation of the paper.

Author Contributes The work done in the project was distributed by Junwei Ma and Jianhua Wang. Junwei Ma has searched the back ground materials. He also designed the analytical characterization and empirical study frame. Jianhua Wang analyzed the data and evaluated the results. All authors have contributed to writing the paper. Junwei Ma has done the critical revision and editing.

Conflicts of Interest The authors declare no conflicts of interest.

References

1. Madeline J, Betsy DG (1999) Predicting cooperative behavior, during a retailer's bankruptcy. Qual Market Res: Int J 2:31–45
2. Stern E-A (1992) Marketing channels. In: Prentice Hall, 4th edn. Upper Saddle River, New Jersey
3. Kin Keysuk, Grazier Gary (1996) Int Mark Rev 13:19–32
4. Zhuang G, Zhou G (2002) Power, conflict and cooperation: an empirical study of channel behavior between Chinese business and industry. Manage World 3:117–124
5. Zhuang G, Li K, Cui X (2008) Relationship marketing orientation and inter organizational relationships impact on corporate relational channel governance. Manage World 7:77–90
6. Liao C, Liu Z (2003) Game Analysis of manufacturers and distributors in channel management. J Chongqing Univ 2:141–144
7. Xiao J (2008) Research on competition and competition between manufacturers and agents in Chinese IT storage market. In: Master thesis of Shanghai Jiao Tong University

The Study on Evaluation Index System of Restructuring Construction Industry Under the Green Development Model

Fang Liu and Jie Zhao

Abstract After the rapid development of construction industry for 20 years in China, the traditional coarse model of the industry characterized by high consumption of the resources, high emission and low efficiency is not sustainable and should be transformed. To build an evaluation index system of economic restructure in construction industry under the green development model is believed to facilitate the successful transformation of construction industry. Based on the theory of green economy and on the analysis of the current situation of the development of the construction enterprises economic transition, the paper analyzes the factors affecting the transformation of construction industry, points out goals and principles for economic restructure in construction industry, establishes primitive evaluation index system. It is intended to assess the extent of transformation so as to realize the sustainable development of construction industry.

Keywords Green development · Construction industry transformation
Index system

1 Current Situations

Green growth for sustainable development has been the main trend in industrialized countries. At the 2015 World Climate Conference, China pledged to reduce its carbon dioxide emissions per unit of GDP by 40–45% by 2020. This is indeed a challenge for a nation where energy mix is dominated by coal and economic growth is heavily dependent on its energy growth. Painstaking efforts must be made to realize this goal and in achieving it, China has to consider how to take initiatives to cope with global climate change.

F. Liu (✉) · J. Zhao
Department of Construction Management, Changshu Institute of Technology, Suzhou, China
e-mail: csfenny@126.com

J. Zhao
e-mail: zhaojie510@sina.com

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_35

As its urbanization accelerates, China has seen outstanding developments in construction industry. The total value of construction output has gone up from 828.2 billion yuan in 1996 to 18,075.7 billion yuan in 2015, a nearly 20-fold increase over two decades [1]. Its growth range far exceeds that of GDP. However, the year of 2015 marks the first decline in the growth rate of construction, which plunged from 10.2% in 2014 down to 2.3% and worries the industry. Faced with the economic downturn and the intricate international environment, the investment structure is bound to change gradually, urging construction enterprises to pace up their economic transition. In this connection, the evaluation index system of economic restructure in construction industry is to play a key role in guiding the construction enterprise transition.

2 Theories Relevant to the Green Development Model and Economic Transition

Traditional industrialization is fundamentally material progress. Yet this highly advanced civilization does not bring about harmony between man and nature as the modern society proceeds; on the contrary, it pushes human civilization away to the opposite extreme. As a new type of civilization that transcends industrialization, eco-civilization explores how harmonious coexistence and coevolution can be achieved between humans and the ecological environment where they live.

Built on eco-civilization, sustainable development addresses the issues along the development between man and nature and among people to balance the relationships between man and nature and among people. As such, the green development model is an inevitable course for both economic ecology progress and sustainable development [2].

2.1 The Sustainable Development Theory

The Blueprint for a Green Economy, published in 1989, first put forward the concept of an "affordable economy" from the perspective of society and from its ecological conditions that economic progress must be affordable for both humanity and environment; the blindness of production growth should be prevented from causing social split and ecological crisis; natural resources should be maintained to sustain economic development. This is an economy that integrates environmental rationality and economic efficiency and will play a dominant role in sustaining economic development.

The impact of sustainable development theory to architectural design is presented in the Guiding Principles of Sustainable Design, released in 2003 in the United States. Its main contents include the notion that the building site should be

selected in line with local ecology while local cultural heritage should be maintained; energy consumption design should take the local climate and environment into consideration by using local building materials instead of non-environmentally friendly ones; during the construction process, great importance should be attached to the environmental pollution and its governance to reduce the pressure on the environment.

2.2 Theory of Corporate Transition

As the concepts of sustainable development and green growth, such as "circulation economy", are gradually put into practice, the number of enterprises that have achieved sustainable and sound development through corporate restructure has increased since the 1980s as theoretical researches on corporate transition have been deepened. Shaheen(1994) argues that transition of enterprises redefines the core competitiveness of enterprises and refers to the changes of concurrent posts, technology, corporate behavior and mode of existence, which brings about pressure on technology against the external environment changes; in this case, the corporate structure should be adjusted accordingly to fully active partial elasticity. The organizational transformation theory proposed by professor Fujikata Seiichi of Division of Business Administration, Meiji University emphasizes that as the life cycle of product and industry changes, the organizational structure and functional departments of the company should be transformed to adapt to the changing environment and in this way the vitality of enterprises can be maintained. In general, those corporate transition theories have laid a theoretical foundation for the evaluation index system of restructure in the construction industry.

2.3 Introduction to the Evaluation Systems of Construction Industry Under the Green Development Mode in Foreign Countries

Some developed countries have conducted research and then developed different standards and evaluation systems for green buildings of various countries, such as Leadership in Energy and Environmental Design (LEED) formulated by U.S. Green Building Council (USGBC), Green Building Tool (GBTOOL) in Canada, Building Research Establishment Environmental Assessment Method (BREEAM) in the United Kingdom, Comprehensive Assessment System for Building Environment Efficiency (CASBEE) in Japan, National Australian Built Environment Rating System (NABERS) in Australia, EcoProfile in Norway and Estadistica de la Calidad Educativa (ESCALE) in France.

3 Analysis of Obstacles in Economic Restructure in Construction Industry

Since industry restructure is a systematic project, prior to the establishment of evaluation index system, attention should be paid to the assessment of major obstacles facing the industry that come from both outside and inside the enterprises. Through on-site researches and analysis of construction companies, major obstacles affecting the economic transition in the construction industry are mainly from the following aspects.

3.1 Internal Environment

The lack of strategic vision of senior management. Many construction enterprises begin to set foot in the real estate industry and take it as an innovative way of corporate transition. In effect, the transition of construction enterprises is more of a strategy while getting involved in real estate is an act of it. Senior management's understanding about economic restructure in the construction industry remains in the expansion of industrial chain and the improvement of anti-risk capability by stepping into various industries.

Short of innovation impetus within construction enterprises. No company can develop without innovation. Yet the technological inadequacy has been an issue in construction industry. Since the beginning of the 21 Century, its technological improvement period has become shorter and shorter due to constant changes of its industry technologies. However, there is a lack of innovation within construction companies, especially that of new construction methods and technologies.

Poor fundamentals of technological innovation. As the construction enterprise improves its internal management, much of the focus has been on the project management of process, safety and cost, rather than on technological innovation. It first results in the lack of professional innovators. Researches show that to foster an innovative corporate culture requires senior management with creative thinking and professional staff with innovative ability. A survey of Dalian construction enterprises conducted in 2010 indicates that technicians only account for 16.9% of the total employees while there are almost no professional technicians to do research and development (R&D) [3]. Fiercely competitive as it is, the construction industry has very poor fundamentals of innovation, partly attributed to unqualified knowledge and skills of practitioners, most of whom have not received good basic education.

Single organizational structure. With the adoption of matrix management mode, conventional construction enterprises have been conducting an extensive operation on the project management. Not only will this single organizational structure lead to an increase of internal consumption and inefficiency of management when the firm reaches a certain scale, but it will result in gaps between research abilities of technicians and technical demand of the project. Thus, impediments are added to the economic restructure of the construction industry in the aspect of organizational structure.

3.2 External Environment

The economic restructure of construction industry is demanded from both internal competiveness within enterprises and external environment—particularly the government. First, in Part X of *the 13th Five-year Plan*, Ecosystems and the Environment, the government has set constraint index for urban air quality, that is to "formulate a plan for ensuring air quality standards in cities are met, strictly enforce obligatory targets, see that cities at and above the prefectural level achieve a 25% reduction in the number of days of heavy air pollution, and channel greater effort into reducing fine particulate matter emissions in key regions. We will establish a monitoring system to ensure that environmental protection standards for vehicles, watercraft, and fuel oil are achieved. We will work to increase the proportion of natural gas users in cities. We will strengthen monitoring of windblown dust from unpaved roads and construction sites" Such government development strategies put tremendous pressure on the construction industry transition.

Second, the construction industry consumes a large amount of energy: the manufacturing, transport and use of cement, steel bar and concrete requires abundant energy during the construction process; a great deal of electricity and thermal energy is expended to keep buildings in operation when they come into service; quite an amount of waste is generated as buildings are demolished, placing massive pressure on the environment while the disposal of them also consumes amounts of energy. According to statistics, the energy consumption in the construction industry in China accounts for 29% of the country's total amount of energy consumption. To fulfill the commitment of capping carbon emission by 2020 in the climate change conference in Copenhagen, it is desperately urgent for China to find a solution to reduce energy consumption in the construction industry and establish reasonable evaluation index system of economic restructure in the construction industry under the green development mode.

4 The Objectives and Principles of Evaluation Index System of Economic Restructure in the Construction Industry

4.1 Transition Objectives

The fundamental goal of the transformation of construction enterprises is not to turn them into high-tech enterprises, but to innovate the original management system within the enterprises, so as to make them meet the requirements of the times and sustainable development. Therefore, with the formulation of the evaluation index system of economic restructure in construction industry, focus should be on transforming them into technological innovation, resource-saving, environment-friendly and eco-friendly enterprises.

4.2 Transition Principles

- Operable

Operability is important to assess the rationality of an index system, including the convenience of accessing materials and data, easy approaches to calculation and possibility of more alternative scientific methods for data that is not readily available.

- Dynamic

The evaluation index system of economic restructure in the construction industry should run through the strategic level of green development in the industry. Since the understanding of green development is gradually improving, priority should be given to the index that can reflect the green growth trend in the construction industry to provide the foundation for policy-making.

- Scientific

Scientific principles can support senior managers to make reasonable and reliable decisions while providing unremitting impetus for developing sound interaction with outside communities and encouraging hi-tech innovation capability.

- Systematic

The evaluation indices are expected to show the economic transition efficiency in the construction industry vertically and horizontally in avoidance of overlapping and vacuum, so as to ensure that the index system is well-organized and systematic. Hence, there should be both static indices that show the current development of construction enterprises and their future potentials, and dynamic indices that indicate the ability of construction enterprises to respond to the external environment changes and their good intention for the environment.

5 Recommended Indices for the Evaluation Index System of Economic Restructure in the Construction Industry

The index for the economic transition in construction industry can be divided into two categories: internal index, referring to basic resource setting, corporate creativity and adaptability to the external environment during the transition of construction enterprises, and external index, which means the benefits for and contributions to the external environment generated during the transition process. For details, please refer to Table 1.

Table 1 Evaluation index system of economic restructure in construction industry

Firs-grade index	Second-grade index	Third-grade index	Evaluation formula
Internal index	Enterprise informationalization rate		The ratio of enterprise information output to investment in enterprise information transformation
	Rate of education background improvement of the management	Basic education background index	The ratio of the education background between the management and staff
		Improved education background index	The proportion of the annual number of managers who receive continuing education to the number of managers
	Rate of R&D application	(1) The annual number of national construction methods	The number of national construction methods per year
		(2) The number of technical improvement team within enterprises	The number of technical improvement team per year
		(3) The growth rate of the number of national and provincial academic institutions	The number of new academic institutions per year/the total number of academic institutions
		(4) The annual number of provincial construction methods	The number of provincial construction methods per year
External index	Carbon emission index	(1) The ratio of annual carbon emission	Carbon emission per year/floor space under construction
		(2) Decrement rate of carbon emission per year	The rate of carbon emission of the year/the rate of carbon emission of the previous year
	Reduction of release of pollutants index	(1) Ratio of sewage discharge fees	Payment of sewage discharge of the year/floor space under construction of the year
		(2) Increment rate of prevention and treatment fees for dust pollution	Payment of prevention and treatment fees for dust pollution of the year/floor space under construction of the year
		(3) The proportion of fees available for on-site security	Payment of on-site security fees of the year/floor space under construction of the year
	Green construction coverage index		The number of green buildings/the total number of construction projects

5.1 Internal Index

- Enterprise Informationalization Rate

This Index mainly shows the improvement level of informationalization within the enterprise and its intention to improve it. The biggest difference between the future construction industry and the traditional one lies in the unprecedented improvement of informationalization in construction industry. Thus, the proportion of enterprise informationalization should be taken as an important index for the construction transition evaluation. This is a dynamic index, whose formula is the ratio of enterprise information output to the investment in enterprise information transformation. A year-on-year increase of the value shows a strong impetus of enterprise informationalization.

- The Rate of Education Background Improvement of the Management

This shows the degree of knowledge upgrade of senior management. For every industry that is in the middle of transition, the willingness of managers, especially the senior managers, reflects in what degree the firm supports its transition. Its evaluation works can take the basic and improved education backgrounds into account. The basic education background is a static index while the improved education background a dynamic one. The basic education background index assesses the education level of original managers, the formula of which is the ratio of the education background between managers and staff. The improved education background index reflects the proportion of the annual number of managers who receive continuing education to the number of managers.

- The Rate of R&D Application

Personnel competence and enterprise informationalization are scientific and technological competitiveness of construction companies during the whole economic restructure. The contribution rate of hi-tech innovation to the transformation of construction enterprises is a core index for the evaluation index system [4]. Its secondary indices include: the annual number of national construction methods, the number of provincial construction methods per year, growth rate of the number of national and provincial academic institutions, the number of technical improvement team within enterprises [5].

5.2 External Index

Under the green development mode, it is of vital importance and major public concern for construction enterprises to minimize the impact on resources and the environment and to properly utilize carrying capacity during their transition [6]. Therefore, great attention should be paid to the environment and resources as to the

establishment of external indices. In this connection, there are three primary indices, namely carbon emission index, reduction of release of pollutants Index and green construction coverage index.

- Carbon Emission Index

Low-carbon construction is one of the fundamental objectives of construction enterprises to achieve economic transition. It is speculated that China will launch the carbon trading market in 2020. When it comes, enterprises that are capable of low carbon construction will win the favor of clients. This index consists of two secondary indices: the ratio of annual carbon emission and the decrement rate of carbon emission per year. The former comes from the ratio of carbon emission per year to floor space under construction while the latter is from the ratio between the rate of carbon emission of the year and that of carbon emission of the previous year.

- Reduction of Release of Pollutants Index

The pollutant control in construction industry is mainly the control of on-site sewage discharge, noise, windblown dust, etc. Since it is hard to determine the amount of pollutant at the current stage the secondary indices are obtained by converting certain external data. The external supervisory authority conducts evaluation on sewage discharge, noise and windblown dust in the construction field to generate relevant data, which can be gained with the help of on-site management platform of construction industry. Then a comparison will be made between these data and the data from the industry. This index comprises three secondary indices: the ratio of sewage discharge fees, increment rate of prevention and treatment fees for dust pollution and the proportion of fees available for on-site security.

- Green Construction Coverage Index

Green construction coverage index refers to the ratio of the number of green buildings to the total number of construction projects. Thereof, different weights should be given to different levels of green buildings and then get the right proportion of green construction within enterprises to corporate operation.

6 Summary and Prospect

As low-carbon economy gains more popularity, it is an inevitable trend for the construction industry to shift from a high energy-consuming business to a green, environment-friendly and low carbon one, which calls for new requirements for both the development of construction industry and its evaluation approach. In this paper, the authors elaborate on the obstacles facing the industry during its economic transition, principles and objectives that should be set with the formulation of the evaluation index system and finally make an analysis of specific indices for the evaluation index system.

However, as this paper deals with the evaluation index system for the economic transition efficiency in construction industry, there is no further exploration of the classification and weights assignment of these indices, which can be a direction in future research.

References

1. Ming C, et al (2015) Research on the relationship between China's total output of construction industry and carbon emissions. Ecol Econ 3:53–56
2. Liu S (2002) Further discussion on eco-civilization and sustainable development. Southeast Acad Res 6:60–66
3. Li H (2010) The study of low-carbon economy systematization mode for construction enterprises. Shandong University of Science and Technology, Qingdao
4. Zhao X (2007) Analysis of obstacles of construction technology innovation in China. Econ Anal 4:46–47
5 Liu Y, et al (2011) Shortcomings and innovative strategies of technology innovation in Chinese construction industry. J Eng Manage 8:359–363
6 Zhao D, Qiu M (2008) Thoughts on sustainable development of construction industry. Econ Res Guide 8:228

Correcting Strategy for Distortion of Ring Part Based on Genetic Algorithm

Ziqiang Zhou, Ke Be and Haoyan Wang

Abstract There are many distorted parts can be used again after correction work. But currently most of the correction is carried out by the workers experience. Workers need extra time of operation to adjust the parameter of correcting equipment. In this paper, a correction strategy combining finite element method (FEM) and intelligent optimization method is proposed for correction ring shaped parts. This method combines the script interface of finite element software Abaqus and the genetic algorithm (GA) program compiled by Python. The process of generating the results of modeling simulation is compiled into a function with adjustable parameters, and then the simulation results are transferred to the operator corresponding to GA. Finally, the optimal correcting parameters is solved. This method overcomes the problem of great difficulty and low accuracy in manual correction parameters, and can be applied into the control systems of automatic correcting equipment.

Keywords Remanufacturing · Correcting method · Genetic algorithm (GA) Control strategy

Z. Zhou (✉)
School of Mechanical Engineering, Changshu Institute of Technology, Changshu, Jiangsu, China
e-mail: zzq_hefei@163.com

Z. Zhou · K. Be · H. Wang
Jiangsu Key Laboratory of Recycling and Reusing Technology for Mechanical and Electronic Products, Changshu Institute of Technology, Changshu, Jiangsu, China

K. Be · H. Wang
School of Mechatronic and Electrical Engineering, China University of Mining and Technology, Xuzhou, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_36

1 Introduction

The remanufacturing or reuse of waste components are getting more and more attention currently. It not only reduce the manufacturing costs but also reduce the resource and energy consumption. In fact, some components can not be used just due to deformation. Since manually correcting of deformed components will increase the cost of remanufacturing, if this work can be carried out with special equipment, the production efficiency can be improved, thus, the cost of remanufactured product is decreased. In this study, the ring shaped parts is focused on. For this type of components, the deformation positions and parameters are different, correcting strategy (including the positions of the supporting points, the positions of the force points, and the amounts of correction displacements) is the key issue for the controller of equipment.

Since the correcting is elastic-plastic process, there are many affecting factors difficult to solve by traditional methods. Currently, correcting parameter are mainly based on approximation methods. However, by this means, the accuracy for parameter is lost. Therefore, the most common used numerical approach, FEM (finite element method), is used for optimizing the correcting parameter of ring shaped part. In the mean time, the optimized algorithm is applied in this work by re-develop programming.

In intelligent optimization algorithm, genetic algorithm (GA) has been widely applied in mechanical manufacturing field. Zhou et al. [1, 2] used a method combining GA and FEM to study the fixture layout optimization. Kulankara et al. [3] used MARC software to combine GA, and optimized the mechanical expanding process of large diameter longitudinal welded pipes, resulting in the optimal process parameters, which can be used in the actual manufacturing. Kulankara and Krishnakumar et al. used GA and finite element method (FEM) to optimize the layout of fixture [4, 5]. Cai et al. applied MARC and the algorithm of GA, and then optimized the mechanical expanding process of large diameter longitudinal welded pipes, resulting in the optimal process parameters [6]. Because there are more uncertainty in the remanufacturing than general manufacturing, the method of parameter optimizing with optimization algorithm will be more valuable.

2 Correcting Method for Distorted Ring Shaped Parts

2.1 Deformation Analysis of Ring Parts

There are various deformation forms of ring shaped parts. In situations of radial tension and compression, the common deformations include elliptical deformation and triangular deformation. In this study, we use a ring part which is subjected to radial compression and deforms into an eclipse as an example to demonstrate the correcting method, as shown in Fig. 1.

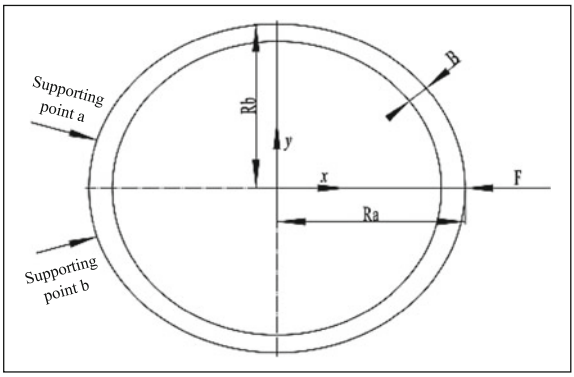


Fig. 1 Correcting of deformed ring shaped part

Firstly, the ring shaped part should be supported at the two radial ends of the parts, with supporting point a and supporting point b. The correction point is at the reverse angle bisector of the connection lines between the two supporting points and the circle center. The purpose for such assignment is to guarantee the stability of the part within the correcting process. At the correcting point, a correction load F is exerted along the direction of the line between the correction point and circle center. With the gradual increase of the correction load, the internal stress of the part gradually exceeds the yield limit of the material, and plastic deformation occur in the part, with exerted correction displacement U. If it is unloaded at this moment, the part will partially rebound under the effect of elastic potential energy, with permanent plastic deformation remaining, that is, the deformed part is corrected.

2.2 Optimizing Model for Correcting Parameters

Optimizing model for correcting parameters is to determine the positions of supporting point, correcting point, and correcting displacements. So that the corrected parts have best correction effect. Therefore, there are 4 design variables, including the positions of the two supporting points, the position of the correcting point, and the correcting displacement. Since the position of the correcting point can be determined by the positions of the two supporting points, this rule can be integrated into the objective functions, and the number of design variables can be reduced to 3. To make the optimization result satisfy the actual requirements, the value ranges of the design variables should be limited. The positions of supporting points are expressed in the form of numbered nodes, i.e., $x_a \in \{0, 1, 2, 3, 4\}$, and $x_b \in \{0, 1, 2, 3, 4\}$, as shown in Fig. 2. The purpose of correction is to recover the deformed parts into their original circular shape, so that the roundness of them should meets the engineering requirements. with FEM, we can establish the function $f(x_a, x_b, y)$ between the roundness of the outer circle of a part, and the design variables of supporting point x_a, supporting point x_b and correcting displacement y. This function is the objective function to be optimized. The constraints

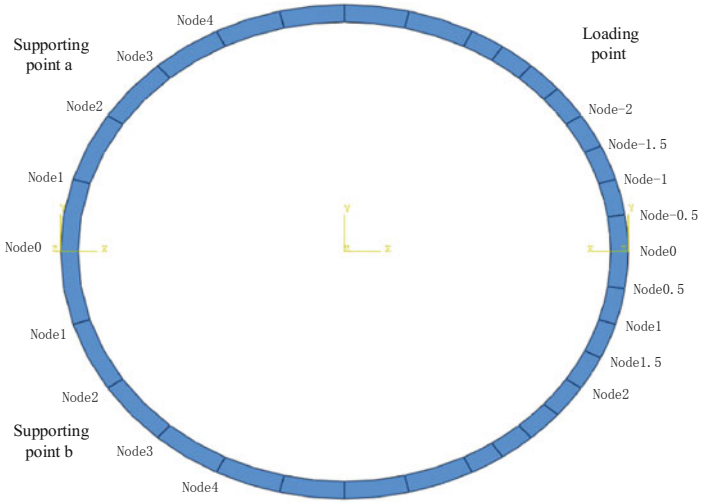


Fig. 2 Boundary conditions and node numbers

of design variables x_a and x_b can be easily determined, i.e., the node numbers of the regulated regions. The constraint of correcting displacement is related to the material property and the initial deformation situation of the parts. In this study, the initial deformation of the part to be corrected is stated in the above, i.e., an approximate eclipse. According to the finite element simulation of the stress along its major axis, the value range of the initial deformation amount for GA optimization algorithm is determined. Then according to the monitoring of maximum stress in FEM, the correcting displacement is constrained. Since the monitoring of maximum stress is added into FEM, the constraint of correction displacement can be initially determined according to users' experience. It should be noted that the constraint value can be set to a large one.

In summary, the optimizing model for ring shaped parts can be defined as follows:

$$\begin{aligned} &\max(f(x_a, x_b, y)) \\ &s.t \\ &x_a \in [0,4],\ x_a \in N^* \\ &x_b \in [0,4],\ x_b \in N^* \\ &\sigma_{\max} = g(x_a,\ x_b,\ y) \le \sigma_b \end{aligned} \tag{1}$$

where

$f(x_a, x_b, y)$ is the optimization objective function;

x_a and x_b are the labels of supporting points a and b;

$\sigma_{\max}$ is the maximum stress within parts in the correction process;

σ_b is the strength limit of the material.

3 GA Algorithm Based Correcting Strategy

3.1 FEM Based Correcting Model

The Fig. 3 is the flowchart of modeling process. Firstly, to facilitate the combination with GA, the pre-processing is defined as a function CreateHuan (Displacement, bound1, bound2, Ra, Rb, B, T). The parameters of the function are the correcting displacement (Displacement, unit m), the node label of supporting point a (bound1), the node label of supporting point b (bound2), the length of the long axis of the approximate ellipse of outer ring (Ra, unit m), the length of the short axis of the approximate ellipse of outer ring (Rb, unit m), the width of the ring part (B, unit m), and the thickness of the ring part (T, unit m).

The flowchart of post-processing is shown in Fig. 4. To Similar the *pre-processing*, function *post-processing* is defined. The usage of *post-processing* function is to call the *pre-processing* function and finish the modeling process. Then it accesses the output database of Abaqus to obtain the deformation information of parts after correcting, and evaluates the deformation information obtained by FEM simulation.

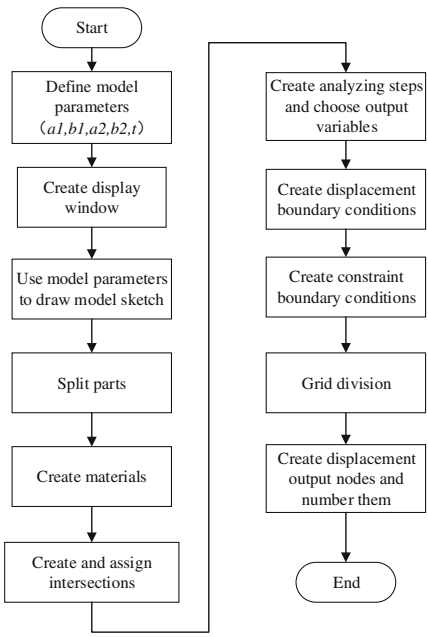


Fig. 3 Flowchart of pre-processing

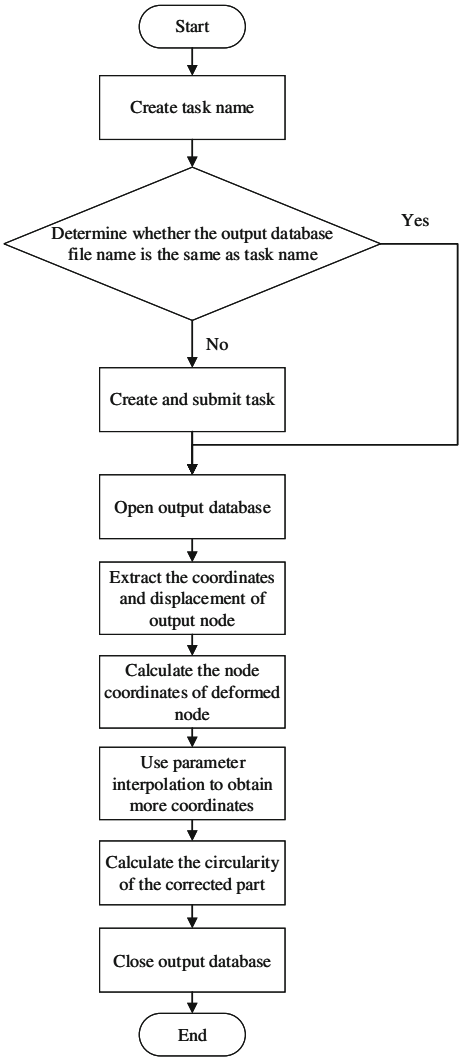


Fig. 4 Flowchart of post-processing

To evaluate the correcting performance, the concept of circularity is introduced. It is a feature parameter used for evaluating the similarity between a contour and a circle, and is defined as follows:

$$c = \frac{A}{r_{\max}^2 \cdot \pi} \times 100\% \tag{2}$$

where A is the area enclosed by the contour; and $r_{\max}$ is the radius of the minimum circumscribed circle of the contour. The value of circularity is within [0, 1], and the circularity of a standard circle is 1. The larger the circularity is, the closer to a circle the contour is.

In this study, the contour is expressed by the coordinates of the observation points of the outer ring of deformed parts. As shown in Fig. 5, 24 observation points are defined at the outer ring, as the output points of FEM simulation. By accessing output database, the coordinates of the observation points after deformation can be extracted. The method is to add the coordinates of the observation points before deformation and the displacement of the observation points.

3.2 DGA Based Optimizing Method

Considering the characteristics of optimization model for correcting parameter, distributed genetic algorithm (DGA) is used to solve the problem. This algorithm divides the population into several sub-populations. Each sub-population not only calculates the individual fitness, but also individually executes selection, cross, and mutation operations, just as each sub-population breeds on its corresponding isolated island. By this means, the optimized program is developed according to the flowchart as shown in Fig. 6. Such similarity check mechanism is suitable for the situations in this study where the number of sub-populations and the number of individuals in each sub-population are both small. Moreover, since the main computational burden of the optimization model in this study focuses on solving FEM, the increment of the computational cost of similarity check can be ignored.

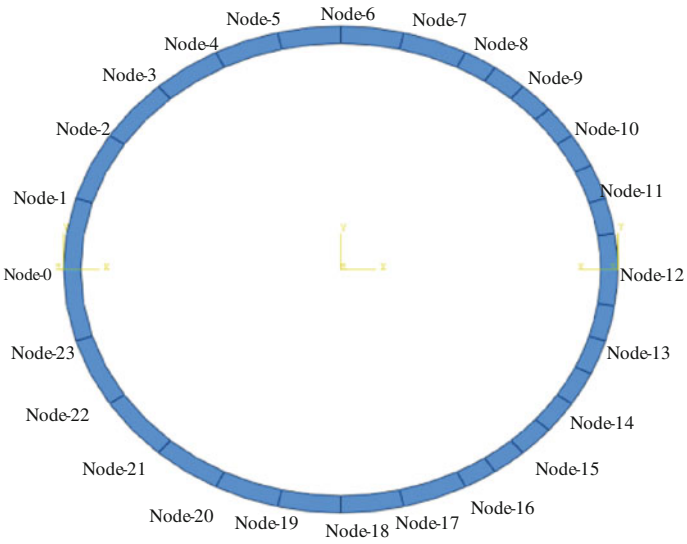


Fig. 5 Labels of output points

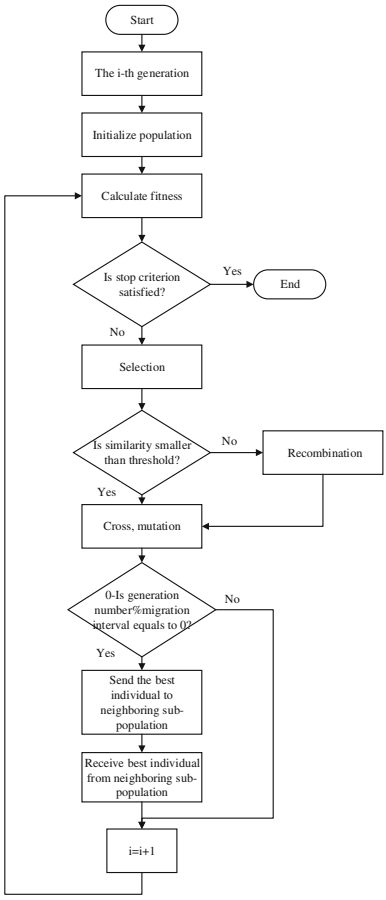


Fig. 6 Program flowchart

4 Case Study

4.1 Modeling for Distorted Ring Shaped Part

In this study, a ring shaped part which subjects to radial stress and deforms into an eclipse is selected for correcting, as shown in Fig. 1. For eclipse shaped deformed parts, due to the specific characteristics of their initial contours, standard eclipse equation is adopted to represent the contour of these parts.

The material of the part is aluminum alloy (7A04), with Young's modulus $E = 73$ GPa, Poisson ratio $v = 0.267$, yield stress $\sigma_s = 571$ MPa, and strengthening factor $k = 0.2$. The shape parameters of the part are listed in Table 1.

In the Abaqus software, it is required that the real stress and real strain be used to define the plastic properties of materials. However, in the uniaxial tension/compression test of materials, data on nominal stress and nominal strain are usually

Table 1 Shape parameters of deformed ring part

Length of long axis Ra (mm)	Length of short axis Rb (mm)	Width B (mm)	Height T (mm)
80	70	5	20

obtained. Therefore, the nominal stress and nominal strain should be converted into real values at first.

Nominal stress σ_{nom} is also called engineering stress, and it is the force per unit area using the area of an undeformed plane, which is given by (3) where F is the force exerted on the plane and A_0 is the area of the undeformed plane.

$$\sigma_{nom} = F/A_0 \tag{3}$$

Thus, nominal strain is the elongation per unit length of undeformed length, $\varepsilon_{nom} = \Delta l/l_0$, where Δl is the elongation and l_0 is the undeformed length.

The relation between real stress and strain and nominal stress and strain is as follows.

$$\begin{aligned} \sigma &= \sigma_{nom}(1 + \varepsilon_{nom}) \\ \varepsilon &= \ln(1 + \varepsilon_{nom}) \end{aligned} \tag{4}$$

where σ and ε are real stress and real strain, respectively.

The real strain ε obtained at this moment contains the total elastic and plastic strain of a material, but in Abaqus, only plastic strain is allowed to define the plastic property of materials. Therefore, the plastic strain should be separated from the total strain, according to the separation equation in (5), where ε^{pl} is the plastic strain, ε^{el} is the elastic strain, and E is Young's modulus.

$$\varepsilon^{pl} = \varepsilon - \varepsilon^{el} = \varepsilon - \sigma/E \tag{5}$$

The nominal stress and strain are firstly converted into real stress and strain according to (4), and plastic strain of the material is separated according to (5), with the results shown in Table 2.

Table 2 Conversion of stress and strain from nominal values to real values

Nominal stress/MPa	Nominal strain	Real stress/MPa	Real strain	Plastic strain
571	0.0078	575.5	0.0078	0
576	0.0184	587.1	0.0183	0.0102
582	0.0291	598.7	0.0286	0.0204
587	0.0397	610.5	0.0389	0.0306
593	0.0503	622.3	0.0491	0.0406
599	0.0609	631.9	0.0592	0.0505
604	0.0716	646.1	0.0691	0.0603
610	0.0822	658.1	0.0790	0.07

4.2 Results and Discussion

It is clear from the above-mentioned model that there are three design variables, namely, the positions of the two supporting points x_a and x_b, and correcting displacement y. The coding of individuals is in the form of list $[x_a, x_b, y]$, where x_a and x_b are integers within [0, 4], and the value of correction displacement y is within 20 mm according to the above stated method. For the simplicity of programming, it is discretized into a value within [0, 20] with a step of 0.5.

The initial population is divided into 3 sub-populations, with each sub-population containing 4 individuals. The individuals of sub-populations are generated randomly, and assigned with initial fitness, with cross and mutation probability being 50%. Functions *CreateHuan* and *postProcessing* in FEM are used to calculate the fitness of individuals of sub-populations. It is worth noting that the output of function *postProcessing* is the circularity c after correction, and scale transform is needed to convert circularity c into the fitness of individuals.

The number of migrated individuals each time is called migration ratio, and it is 1 in this study. Migration interval refers to the time interval between two migrations, and it is set to 4 generation in this study.

The convergence situation of correction parameter optimization is shown in Fig. 7, where the vertical axis represents the average circularity of a generation, and the horizontal axis represents the generation number. It is clear in the figure that with the proceeding of optimization, the circularity of the parts after correction constantly increases.

It is clear that the convergence speed of the selected numerical example is fast. Since the convergence form is selected as constrained objective function value, the generation number of convergence of the optimization function is not fixed. In this study, the generation number of convergence for the selected numerical example is within 20.

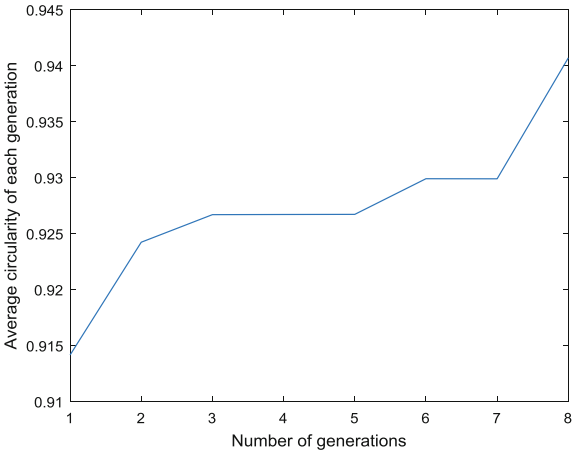


Fig. 7 Convergence process of correcting parameter optimization

The result of optimal model is solved as follows. The positions of supporting points a and b are both at the node with number 0. In other words, the supporting and correction are along the direction of the major axis of the eclipse. The correction displacement is 15 mm, with the initial circularity and corrected circularity of the ring part being 0.861 and 0.942, respectively. The displacement cloud chart of the corrected ring shaped part is shown in Fig. 8. The coordinates of the observation points before and after correction is shown in Fig. 9.

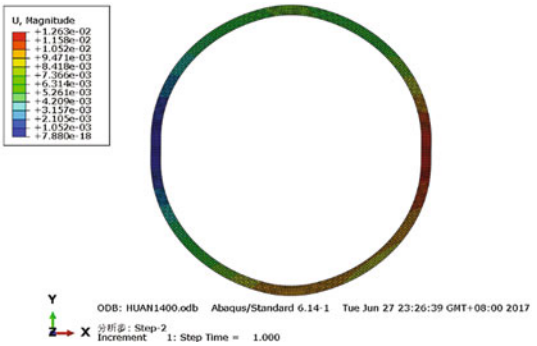


Fig. 8 Part deformation under optimal correcting parameters

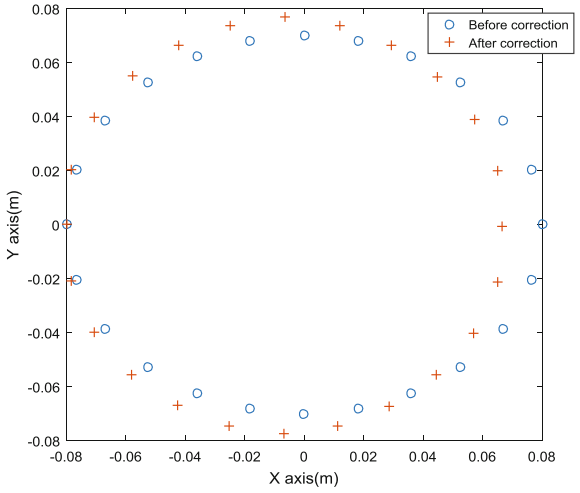


Fig. 9 Coordinates of observation points before and after deformation

5 Conclusions

In this study, a correcting strategy for ring shaped distorted parts is proposed based on FEM and GA. A deformed ring part with eclipse shape is used as an example to demonstrate the secondary development method for the script interface of FEM software Abaqus, and the application of improved GA in correction optimization model. It can be concluded from the results that the method can accurately find the optimal correction parameters, solving the large difficulty and low accuracy problems of numerical solution, and this method can be applied in the control systems of automatic correction equipment.

References

1. Zhou X, Zhang W, Qin G et al (2005) On optimizing fixture layout and camping force simultaneously using genetic algorithm. Mech Sci Technol 24(3):339–342 (In Chinese)
2. Krishnakumar K, Melkot SN (2000) Machining fixture layout optimization using the genetic algorithm. Int J Mach Tools Manuf 40(4):579–598
3. Kulankara K, Satyanarayana S, Melkote SN (2002) Iterative fixture layout and clamping force optimization using the genetic algorithm. J Manuf Sci Eng 124(1):119–125
4. Ziqiang Z, Guohong D, Xiangyan Z, Yongjian Z (2015) Remanufacturing strategies based on value analysis of product life cycle. Open Cybern Systemics J 9:2826–2833
5. Liu Y, Xu B, Shi P, Liu B (2011) Assessment indexed of used products remanufacturability. China Surf Eng 24(5):94–99 (In Chinese)
6. Ramesh S, Huisingh D, Chinnam RB, Subramoniam S (2013) Remanufacturing decision-making framework (RDMF): research validation using the analytical hierarchical process. J Clean Prod 40(9):212–220

Human Centered Automation System for ELV Disassembly Line

Guohong Dai and Ziqiang Zhou

Abstract Since there are so many types of end of life vehicle need to be worked on the disassembly line. Especially, there parameters and the state are different, thus, it is difficult to use generally termed automation for disassembly work. In this paper, the concept of human centered automation is explained and applied on the field of disassembly line for end of life vehicle. Firstly, the state information of end of life vehicle is acquired by the human expert. And then, with RFID technology and expert system, the operating information is created and displayed on the screen of each work station of the disassembly line. By this means, the disassembly work can be finished by the workers according to the basic information. The detailed disassembly operation still finished based on the experience and knowledge of workers. In the end, a case study is given and discussed.

Keywords Disassembly · End of life products · Human centered automation RFID

1 Introduction

Because there are so many brand, model type and parameters of ELV (end of life vehicle), it is difficult to modelling all the operating information with uniform data structure. The actual disassembly work is carried out by the workers according to their experience. If mass disassembly work for ELV is lie on the works experience, the production efficiency of disassembly can not be guaranteed. Therefore, with

G. Dai
Changzhou University, Changzhou 213164, Jiangsu, China

Z. Zhou (✉)
Changshu Institute of Technology, Changshu 215500, Jiangsu, China
e-mail: zzq_hefei@163.com

Z. Zhou
Jiangsu Key Laboratory of Recycling and Reusing Technology for Mechanical and Electronic Products, Changshu Institute of Technology, Changshu 215500, Jiangsu, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_37

information system and automation system to realize a disassembly process planning platform, it will be help for improving the degree of standardization, and it will be help for increasing the rate of reuse of component after disassembly.

Recycling of ELV has got much more attention from the world, both developed countries and developing countries. In the European countries, according to the high labour cost, most of the ELV are disposed with shredding technology. And then, the valuable material is sorted from the shredding material [1]. While in China and several countries, in order to improve the recycling rate, the ELV can be treated with multi-goal disassembly method. That is, in the process of disassembly, some of the components are disassembled for remanufacturing, and some of the components are disassembled for material recycling [2–4]. For the purpose of increasing efficiency of disassembly, several vehicle manufacturer of European countries, including Volkswagen, Skoda Octavia, BMW etc., developed an information system which named IDIS (International Dismantling Information System) for sharing disassembly information on the Internet [5]. Because the increasing rate of private car in China is so higher than other counties, the new technology for ELV with high efficiency and high recycling rate is needed.

From the view of disassembly line, nearly most of the research work are focused on the model of disassembly or off-line planning for disassembly line. The real time control system for disassembly line is not mentioned before. Yet, it is difficult to construct a real time automation system for ELV disassembly line since there are so many uncertainty factor in the disassembly work. Before decades, a new concept of automation, human centered automation, is proposed for human cooperated work [6]. Its idea is to use automation technology to enhance the capabilities of human workers, and responsible for the safety and effectiveness of complex dynamic systems. This approach will be play a great role in the filed of ELV disassembly.

2 Disassembly Operation Mode of End-of-Life Vehicles

Currently, there are two type of disassembly mode for ELV, one is disassembly line method (illustrated in Fig. 1), and another one is multi-point disassembly method (illustrated in Fig. 2). For the disassembly line, there are many processing equipment located along the conveyer line. Since the operating time for each ELV is not the same, that the scheduling system is very important for whole disassembly line.

For the Fig. 2 showed multi-point disassembly method, only several lifting and turning equipment is needed for ELV disassembly. That is named disassembly unit in the Fig. 2 after disassembly work is finished, a forklift is used to transfer ELV from disassembly to the baler equipment. It is clear that, in the method, the efficiency of disassembly is not the same with the method of disassembly line.

According to the concept of human centered automation, no matter in the disassembly line method or multi-point disassembly mode, the workers skill and experience is very important. But because there are so many uncertainty in the disassembly process for ELV. The real time scheduling system is needed in order to improve the efficiency of disassembly work and the recycling rate of component and the material.

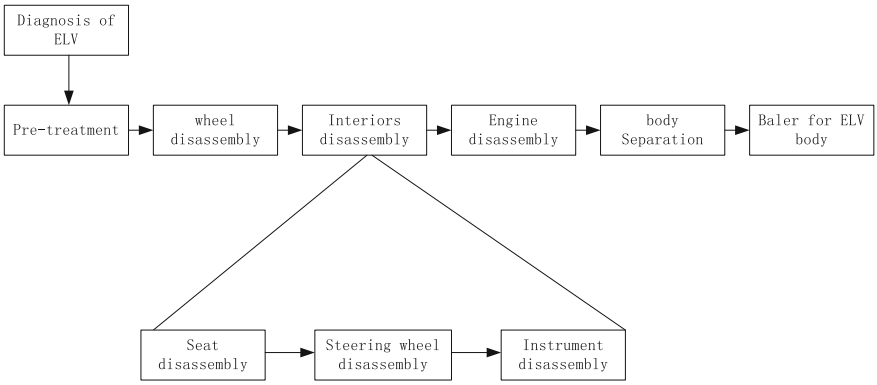


Fig. 1 Principle of the disassembly line mode

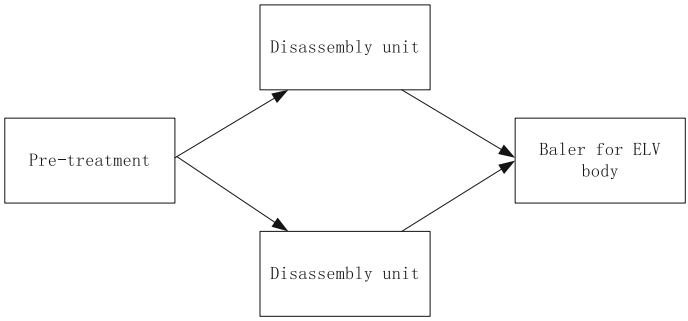


Fig. 2 Principle of multi-point disassembly method

3 Hardware Structure of ELV Disassembly Line Scheduling System

Hardware structure of ELV disassembly computer aided process planning system is shown in Fig. 3. The whole system work on the enterprise computer network platform, and also maintains the contact with the internet at the same time. The data service software of the system works on the network server, and terminal data interaction software is installed on each workstation computer of the disassembly line, in order to display the process information of the station. These terminal computers are linked to the server through the enterprise network. Before the dismantling of ELV, the diagnostic technician will enter the vehicle diagnostic information into the server by wireless mode through the handheld computer (Tablet PC), also generate a serial number and store it on the RFID (Radio Frequency Identification Devices) chip, then attach the chip to the designated position of the ELV. In the disassembling process of ELV, when the ELV on the tray enters a dismantling station, firstly, the serial number information is read

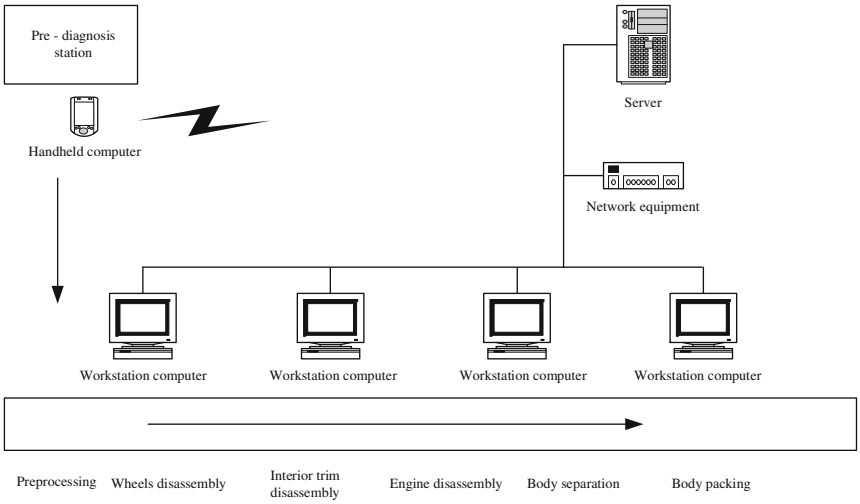


Fig. 3 Hardware structure of ELV disassembly computer aided process planning system

through the RFID reader of the station, then the corresponding disassembly information is obtained from the server according to the serial number information, which includes disassembling tools, disassembling methods, disassembling parameters (the relevant auxiliary equipment needs adjusting parameters, such as pressure, voltage, etc.), and the information is displayed on LED screen which is above the station, so it's easy for workers to check. The dismantling tool is prepared by the tool management which can be distributed or self-contained as required. When all the disassembly work is completed, the body of the ELV needs to be packaged. In this station, the workers will take down the RFID chip which can be used repeatedly, and then start the packaging program, after the disassembly detection, the reusable and remanufactured parts will be sent into the warehouse, meanwhile the information of the receipted parts will be entered into server, and when the disassembly parts are taken out, the flow information will be recorded.

4 Software Structure of ELV Disassembly Line Scheduling System

The software structure of ELV disassembly computer aided process planning system is shown in Fig. 4. The main part of the system is the knowledge base section, which include disassembly rule management, vehicle model database and disassembly process planning module, in which the vehicle model database mainly records and stores the parts parameter information table of the vehicle. Disassembly process planning module is designed and implemented according to the reasoning mechanism of generative rules, and using the prior definition of partial destruction disassembly rules to determine the way and the tools for disassembling the main

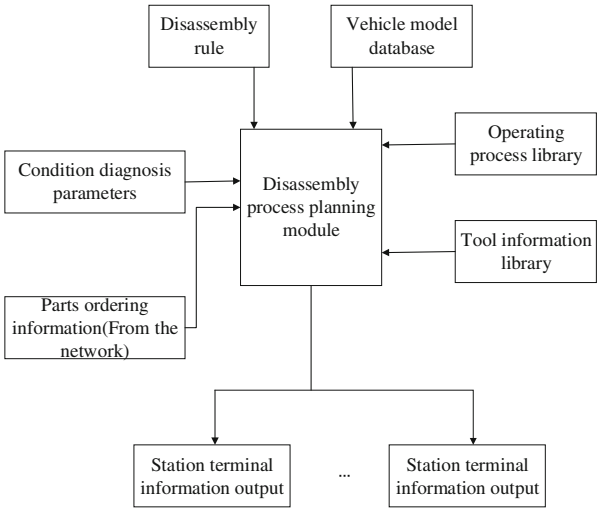


Fig. 4 The software structure of ELV disassembly computer aided process planning system

parts. The condition diagnosis module is used to record the condition of the diagnostic parameters and put into the system. In addition, parts ordering information is the data interface associated with the Internet, it is used to input the demand information of some special parts for the system to process. Such as the hood generally processed as a material recovery process, which means it can be operated by destructive disassembly. However, if there is an external demand order, the brand vehicle's hood needs to be disassembled by the no damage way. This operation information needs to be output in time and displayed on the operating station of the disassembly line to determine the way and the tools for disassembling the main parts. The condition diagnosis module is used to record the condition of the diagnostic parameters and put into the system. In addition, parts ordering information is the data interface associated with the Internet, it is used to input the demand information of some special parts for the system to process. Such as the hood generally processed as a material recovery process, which means it can be operated by destructive disassembly. However, if there is an external demand order, the brand vehicle's hood needs to be disassembled by the no damage way. This operation information needs to be output in time and displayed on the operating station of the disassembly line.

4.1 Data Model for Disassembly Process

The basic structure of the dismantling process data model is shown in Fig. 5. It is mainly composed of basic information, disassembling information and process information of disassembly process. The basic information mainly includes the car's parts structure information, and related attributes. In addition, the parameter information of the vehicle is also included. The disassembly information mainly

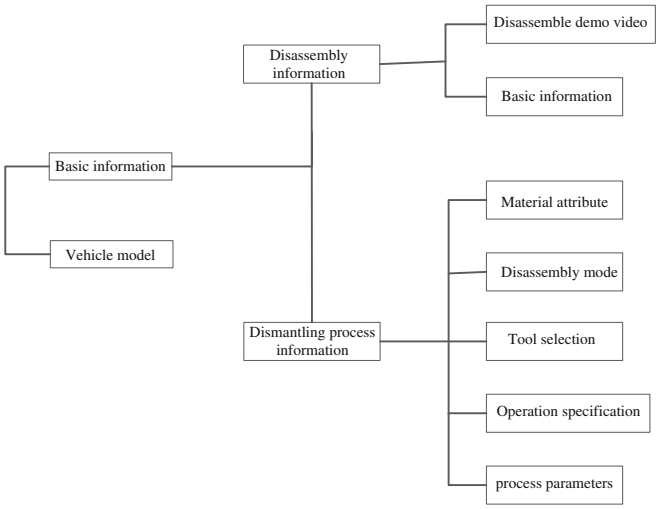


Fig. 5 Process data structure model of disassembly line

describes the position of the parts to be processed in the vehicle and the key operation position when deal with the parts by the pictures. Standard operating method for disassembly of components is demonstrated by the disassembling demonstration video. After the configuration and output through the station terminal, the operator is provided as the operation instruction The process information of the dismantling process mainly describes the material information of the dismantling object, what kind of dismantling mode is adopted, what type of tool is used, the specific operation steps, the demolition path and the setting of the key process parameters of some equipment.

The function of the dismantling process data management module is to complete the editing of the vehicle dismantling process. The technician uploads each part of the disassembly resources of a certain type of vehicle, including: documents, pictures and videos, and then edit and manage the disassembly information of the new model.

4.2 Disassembly Job Planning Based on Rule Reasoning

In order to carry out computer-aided management and decision-making of the disassembly process. This paper gives the basic disassembling strategy of vehicle type by the process reasoning module based on production rule. The antecedent of a production rule consists of regular objects, attributes, and logical relationships, the rule consequence is mainly composed of the disassembly method or the type of rule that is further processed. The rule object refers to the parts that need to be classified and processed in the process of disassembly, such as doors, engines, lights, etc., and the parts that do not need to be classified are not the objects in the antecedent of the rule. Figure 6 is the program interface of the rule editor.

Fig. 6 Disassembly rule editor

4.3 *Information Management and Scheduling of Disassembly Station*

In order to facilitate the disassembly of online workers to obtain information, this study through the large-screen display to output system planning and access to dismantling information. The disassembly information is consist of video, graphics and text. Workers can read the dismantling information which has classified and have the further view to the relevant information. The information output interface of the station terminal is shown in Fig. 7.

Fig. 7 Disassembly information output of the disassembly line station terminal

Tool management and scheduling module mainly includes tool scheduling, tool planning management and so on. The tool scheduling means send the preparation information of the tool according to the current disassembling task of each station, and the borrowing management of the tool, then the borrowing management of the tool is carried out. Record and manage the borrower and the borrowing date of the tool. Its functional relationship is shown in Fig. 8.

The information management of the disassembly tool is shown in Fig. 9. In this module, you can view and edit the information of the tool, including the tool name, tool number, number of tools, tool type, and tool use maintenance information.

In the tool matching module, select the vehicle type, parts, disassembly tools and their number to complete the matching of tools, when multiple tools are required to match, they should be arranged in the order of operation.

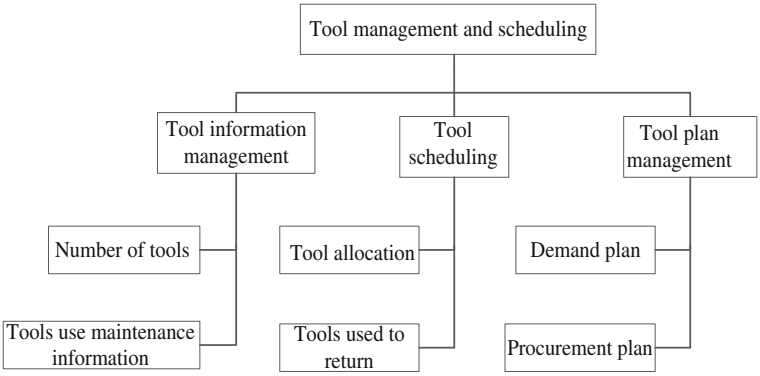


Fig. 8 Tool management and scheduling function module diagram

Fig. 9 Information management of disassembly tools

5 Conclusion

Due to the large number of end-to-life vehicles, the operation of the disassembly process contain a lot of information. The design of a fully intelligent disassembly process planning system is unrealistic and unnecessary. From the actual situation of the current end-to-life vehicles disassembly enterprise, the main hope is to solve the information management of the disassembly process, and provide prompt help for the key steps of the disassembly process. Therefore, it is reasonable to make full use of the information management function and retrieval function of computer network to complete the process planning and process management of end-to-life vehicles disassembly line.

Acknowledgements This work was financially supported by the Jiangsu Science & Technology Pillar Program of China (No. BE2013060).

References

1. Kanari N, Pineau J-L End-of-life vehicle recycling in the European Union. www.tms.org/pubs/journals/jom/0308/kanari-0308.html
2. Chen M (2005) End-of-life vehicle recycling in China: now and the future. JOM J Miner 57 (10):20–26
3. Zamudio-Ramirez P The automobile recycling industry in North America. www.systemdynamics.org/conferences/1999/PAPERS/PARA196.PDF
4. Fuse M, Kashima S (2008) Evaluation method of automobile recycling systems for Asia considering international material cycles: application to Japan and Thailand. J Mater Cycles Waste Manage 10(9):153–164
5. Ferrao P, Amaral J (2005) Assessing the economics of auto recycling activities in relation to European Union directive on end of life vehicles. Technol Forecast Soc Chang 73(3):277–289
6. https://link.springer.com/chapter/10.1007%2F978-1-4613-1447-9_31

Rolling Bearing Fault Diagnosis Using Deep Learning Network

Shenghao Tang, Yuqiu Yuan, Li Lu, Shuang Li, Changqing Shen and Zhongkui Zhu

Abstract Automatic and accurate fault diagnosis of rolling bearing is crucial in rotating machinery. Deep belief network (DBN) can automatically learn valid features from signals, which leaves out manual feature selection compared with traditional fault diagnosis methods. In this paper, a novel method called deep belief network with Nesterov momentum is developed for the diagnosis of rolling bearings. Nesterov momentum is used to accelerate training and improve precision. An experimental analysis is carried out using a dataset under different bearing health states from a test rig to substantiate the utility of the proposed DBN architecture. Results show that the method demonstrates impressive performance in bearing fault pattern recognition. Comparison analyses are further conducted to demonstrate that Nesterov momentum can improve the capability of DBN.

Keywords Fault diagnosis · Feature learning · Nesterov momentum
Deep belief network

1 Introduction

Rolling bearings are essential parts that are widely used in rotating machinery. With the rapid development of national industry, rotating machinery always runs under complex operating conditions like heavy load, high temperature and high speed. Rolling bearings will inevitably suffer fault, which results in severe machine breakdowns and yields serious economic losses and human casualties. Therefore, it is of great importance to accurately and automatically identify rolling bearing faults. Bearing fault diagnosis has been a hot research topic in recent decades [1].

Numerous methods have been applied successfully to bearing fault diagnosis, such as empirical mode decomposition (EMD), wavelet packet transform (WPT),

S. Tang · Y. Yuan · L. Lu · S. Li · C. Shen · Z. Zhu (✉)
School of Urban Rail Transportation, Soochow University,
Suzhou 215131, Jiangsu, People's Republic of China
e-mail: zhuzhongkui@suda.edu.cn

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_38

and support vector machines (SVMs) [2–4]. Although the aforementioned methods have been reported in machine fault diagnosis, they still have several shortcomings. For EMD and WPT, extracting features manually requires a high degree of expertise in signal processing. For SVM, the diagnosis performance of model depends heavily on the quality of the selected features which largely depend on prior knowledge of diagnosticians.

Deep belief network [5] has lately gained popularity as a new method in machine learning because of its capability to extract effective features automatically in fault diagnosis. Tran et al. [6] employed DBN and Teager–Kaiser energy operator to recognize the fault conditions of compressor valves. Gan et al. [7] combined DBN with WPT to establish a hierarchical diagnosis network for complex fault states recognition of rolling bearings. However, most existing methods still require artificial feature extraction from data or complex signal processing for pre-training. In addition, most of current work added momentum to the original DBN to increase the learning speed. However, momentum may probably miss the local optimum, leading a poor identification performance. Therefore, Nesterov momentum (NM) is used to replace momentum in this study. In the NM-DBN, frequency domain signals are fed for fault diagnosis. The advantages of the proposed method are as follows: (1) the use of frequency domain signals without complex signal processing, (2) the capability of extracting deep fault features automatically and sensitively, and (3) a better identification performance compared with momentum-DBN.

The reminder of this paper is arranged as follows. Section 2 presents the basic theories of DBN. Section 3 details the NM algorithm. Section 4 confirms the utility of the proposed DBN model in bearing fault diagnosis and compares it with momentum-DBN and original DBN. Section 5 presents the conclusions.

2 The Basic Theories of DBN

2.1 Restricted Boltzmann Machine (RBM) Architecture

RBM is regarded as a Markov random field. As is shown in Fig. 1, each RBM is composed of a visible layer with m visible neurons $v = \{v_1, v_2, \ldots, v_i\}$, and a hidden layer with n hidden neurons $h = \{h_1, h_2, \ldots, h_j\}$. The neurons of visible layer are fully connected to the neurons in hidden layer by the weight vector W. The vectors a and b represent the biases in visible and hidden layer, respectively. The energy function is defined as:

$$E(v, h; \theta) = -\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij} v_i h_j - \sum_{i=1}^{m} a_i v_i - \sum_{j=1}^{n} b_j h_j \tag{1}$$

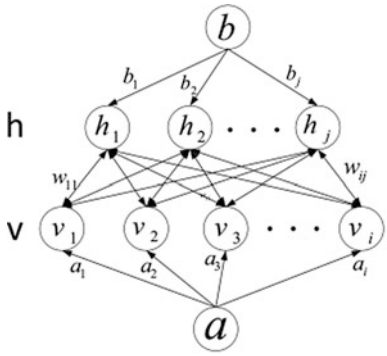


Fig. 1 Schematic architecture of an RBM

where $\theta = \{W, a, b\}$, v_i is the status of the ith visible neuron, and h_j is the status of the jth hidden neuron. The joint probability distribution to pairs of visible and hidden vectors can be calculated by Eq. (2):

$$P(v, h; \theta) = \frac{1}{Z(\theta)} \exp(-E(v, h; \theta)) \tag{2}$$

where $Z(\theta)$ is a normalization factor defined by $\sum_v \sum_h \exp(-E(v, h; \theta))$.

No two neurons in a one same layer are connected, which leads that visible and hidden neurons are conditionally independent. Thus, the conditional probabilities are defined by Eq. (3):

$$\begin{aligned} P(h|v; \theta) &= P(v, h; \theta)/P(v; \theta) = \prod_j P(h_j|v; \theta) \\ P(v|h; \theta) &= P(v, h; \theta)/P(h; \theta) = \prod_i P(v_i|h; \theta) \end{aligned} \tag{3}$$

The individual activation probabilities can be expressed as follows when considering binary neurons:

$$\begin{aligned} P(h_j = 1|v; \theta) &= \sigma\left(\sum_{i=1}^{m} w_{ij} v_i + b_j\right) \\ P(v_i = 1|h; \theta) &= \sigma\left(\sum_{j=1}^{n} w_{ij} h_j + a_i\right) \end{aligned} \tag{4}$$

where $\sigma = 1/(1 + e^{-x})$ is a sigmoid function. Based on log-likelihood function, parameters are solved by the following update rule:

$$\begin{aligned} \partial \ln p(v; \theta)/\partial w_{ij} &= \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{\text{model}} \\ \partial \ln p(v; \theta)/\partial a &= \langle v_i \rangle_{data} - \langle v_i \rangle_{\text{model}} \\ \partial \ln p(v; \theta)/\partial b &= \langle h_j \rangle_{data} - \langle h_j \rangle_{\text{model}} \end{aligned} \tag{5}$$

where $\langle\cdot\rangle_{data}$ represents data distribution expectation and $\langle\cdot\rangle_{\text{model}}$ represents the model-defined distribution expectation. An unbiased sample of $\langle\cdot\rangle_{data}$ is easily gained but the sample of $\langle\cdot\rangle_{\text{model}}$ is unsure. Therefore, contrastive divergence is used to approximate gradient by one full-step of Gibbs sampling. The updated rule is modified as follows:

$$\begin{aligned}\Delta w_{ij} &= \eta_w(\langle v_i h_j\rangle^0 - \langle v_i h_j\rangle^1)\\ \Delta a &= \eta_a(\langle v_i\rangle^0 - \langle v_i\rangle^1)\\ \Delta b &= \eta_b(\langle h_j\rangle^0 - \langle h_j\rangle^1)\end{aligned} \tag{6}$$

where η_w, η_a, and η_b respectively denote the learning rates of weight, visible and hidden bias, which range from 0 to 1.

2.2 DBN Theory

DBN is a deep learning model constructed by stacking several RBMs layer by layer. The architecture of a DBN with three stacked RBMs is shown in Fig. 2. Layer 1 (visible layer v) and layer 2 (hidden layer h_1) compose the first RBM, layer 2 (hidden layer h_1) and layer 3 (hidden layer h_2) compose the second RBM, and the third RBM is formed by layer 3 (hidden layer h_2) and layer 4 (hidden layer h_3).

The DBN classification process can be generalized into two steps, namely, pre-training and fine-tuning. In the pre-training, the data are fed into the first RBM to learn the weight and bias, then the features learned by the trained RBM are used as input to the following RBM. Each RBM is trained by features from the previous RBM as its input data. Subsequently, the features learned automatically by DBN that connected by these trained RBM layers are fed into a softmax classifier. In the fine-tuning, back propagation algorithm is employed for improving the performance of the model.

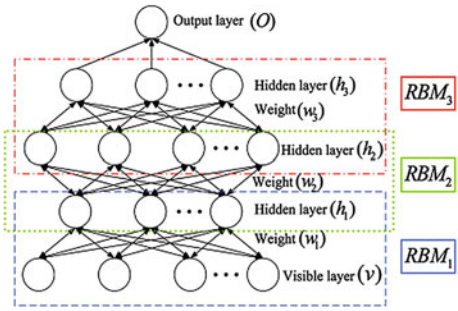


Fig. 2 Schematic architecture of a DBN

3 Theory of the Proposed Model

3.1 Nesterov Momentum

Mini-batch stochastic gradient descent (SGD) is generally used for gradient descent on the likelihood function. However, mini-batch SGD is easily trapped in ravines and oscillates across the slopes of the ravines and makes uncertain progress toward the local optimum, which leads to slow training speed. Momentum [8] is always applied to accelerate mini-batch SGD through the ravines and dampen oscillations:

$$\begin{aligned} v_t &= \gamma v_{t-1} + \eta \nabla_\theta J(\theta) \\ \theta &= \theta - v_t \end{aligned} \tag{7}$$

where γ is a momentum factor (generally set to 0.9 or an approximate value), v_t denotes the current update vector, which is formed by a proportion γ of the previous update vector and the present gradient. When the dimensions of the momentum term and the gradient are in the same directions, movement in this direction is accelerated. Thus, faster convergence is achieved. If the dimensions of the momentum term and the gradient are in the opposite directions, then oscillation is reduced.

However, momentum is unable to determine where the next step is going but continues accelerating gradient descent, which may miss the local minimum when the slope starts to change from falling to rising. NM [9] is an effective method to provide the momentum term with prior knowledge. An approximation of the next position of the parameter θ can be obtained by computing the term $\theta - \gamma v_{t-1}$, which helps momentum term roughly judge where the parameters are proceeding, and effective decline in advance is obtained.

$$\begin{aligned} v_t &= \gamma v_{t-1} + \eta \nabla_\theta J(\theta - \gamma v_{t-1}) \\ \theta &= \theta - v_t \end{aligned} \tag{8}$$

3.2 NM-DBM Fault Diagnosis

The flowchart of learning process of the NM-DBN model is presented in Fig. 3. In this study, frequency domain signals are selected as the input of the model and are normalized during signal preprocessing because the value of the sigmoid activation function is limited within $[0, 1]$.

In the pre-training, the *CD*-1 algorithm is adopted to obtain gradient, which is utilized to update the weights and biases in NM. The output from the previous RBM is used to train the next RBM. The output of the last RBM is set as the features learned from the training dataset. In the fine-tuning, a softmax classifier is

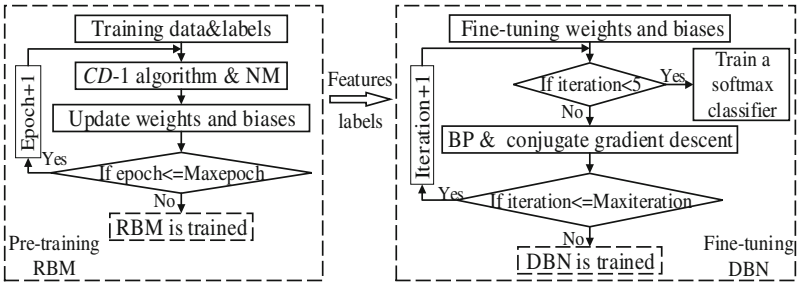


Fig. 3 Learning process of the NM-DBN model

trained separately, in which features with training labels are fed in the initial iterations. The training dataset with labels is used as the input of the entire model for the optimal parameters of the model through back-propagation (BP) algorithm and conjugate gradient descent. The error is defined as the deviation of the classifier output and the target label.

4 Experimental Validation

4.1 Experimental Data-Sets

The defective bearing dataset used for experimental validation are acquired through an accelerometer installed near the testing bearing. As illustrated in Fig. 4, the experimental setup consists of a drive motor, a healthy bearing and a testing bearing, which supports the motor shaft. The type of testing bearing is 6205-2RS SKF, and a simple device based on thread-and-nut system is assembled in the radial direction to add an adjustable mechanical load. The testing bearing inner race (IR), outer race (OR) and roller defects are artificially set with a width of 0.02 mm by a wire-electrode cutting machine. The sampling frequency is 10 kHz and the rotating speed is 961 rpm.

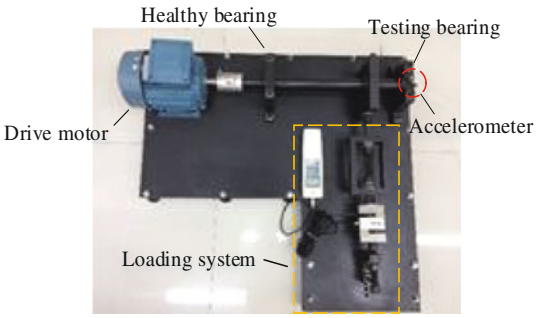


Fig. 4 Experimental setup for obtaining the vibration signals of the rolling bearings

A total of 400 samples for each state are obtained where 300 samples are used for training and the others are for testing. The training and testing procedures are implemented ten times for validation. The number of sampling points in a period is 624, and thus a time domain sample with 1248 points is regarded to contain sufficient information that reflects the health states of the bearing. The related frequency domain sample contains 624 points accordingly. Labels 1, 2, 3 and 4 represent health, OR fault, IR fault and ball fault for a softmax classifier-based DBN model, respectively.

The NM-DBN model consists of one visible layer with 624 neurons and two hidden layers; one of the hidden layers has 600 neurons and the other has 200 neurons. The initial learning rates of two RBMs are set as 0.006 and 0.001.

To demonstrate the advantages of the proposed scheme, the comparative methods are investigated: (1) original DBN, (2) momentum-DBN. In original DBN, mini-batch SGD is used directly to gradient descent. These methods are performed ten times as well.

4.2 Results and Discussion

The results of the three method are listed in Table 1. The average training accuracy and testing accuracy of the proposed method are 99.98% and 99.95%. The NM-DBN model obtains the highest training and testing accuracy than the other two models. However, the Momentum-DBN achieves a lower training and testing accuracy than original DBN, which means that the Momentum-DBN may miss the local optimum.

Figure 5 shows the convergence speed of training error in the pre-training process. The original DBN model achieves the slowest convergence in both first and second RBMs training. The NM-DBN achieves better convergence performance than original DBN, which converges in a manner similar to momentum-DBN. Hence, Nesterov momentum and momentum methods can effectively accelerate the pre-training process.

Training errors and accuracies of the three models in fine-tuning process are shown in Fig. 6, the NM-DBN model and momentum-DBN model have similar initial values of training errors which are smaller than that of original DBN, and their errors drop at a similar velocity in the initial iteration, which is due to the acceleration in the pre-training process. However, momentum-DBN misses the local optimum in the subsequent iterations. As a result, the speed becomes

Table 1 Diagnosis results of the three methods

Method	Average training accuracy (%)	Average testing accuracy (%)
NM-DBN	99.98	99.95
Momentum-DBN	99.76	99.38
Original DNM	99.96	99.88

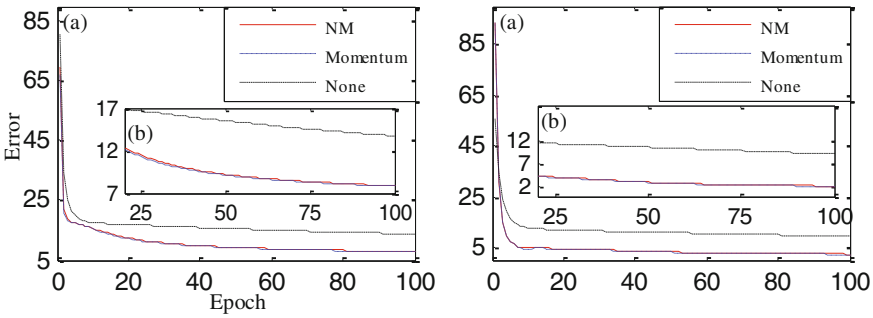


Fig. 5 **a** Training errors of the two RBMs and **b** zoomed plot

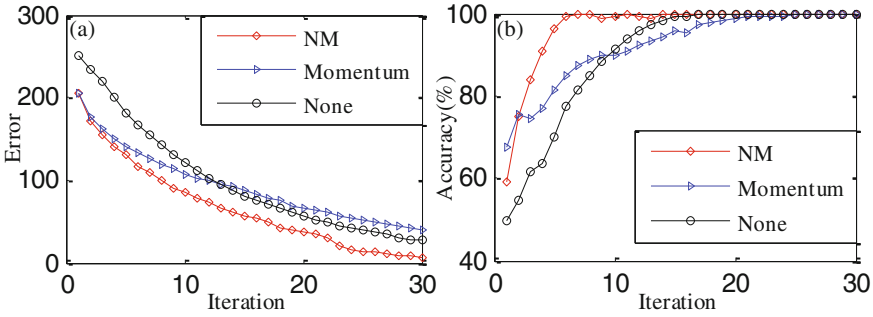


Fig. 6 **a** Training errors and **b** accuracies of the three models

subsequently slow, and momentum-DBN model obtains worse performance than original DBN model. In contrast, the NM-DBN exhibits its advantages in terms of training error and dropping speed of training error, compared with the other models. Consequently, it achieves the best performance of fault diagnosis among these methods.

5 Conclusion

In this study, a novel NM-DBN model is presented and applied successfully to bearing fault diagnosis. In the view of the disadvantage of momentum, Nesterov momentum is employed to replace momentum to design an NM-DBN model. The data of the defective bearing is used to evaluate the effectiveness of the proposed NM-DBN method. The NM-DBN model can offer automatic representative feature extraction to avoid complex manual feature extraction from frequency domain signals. The proposed DBN demonstrates better classification performance on the defective bearing dataset compared with momentum-DBN and original DBN models.

Acknowledgements This research was financially supported by the National Natural Science Foundation of China (Grant No. 51505311 and 51375322), the Natural Science Foundation of Jiangsu Province (No. BK20150339) and the China Postdoctoral Science Foundation funded project (2016T90490).

References

1. Lu S, He Q, Kong F (2014) Stochastic resonance with Woods-Saxon potential for rolling element bearing fault diagnosis. Mech Syst Signal Process 45:488–503
2. Guo W, Tse PW, Djordjevich A (2012) Faulty bearing signal recovery from large noise using a hybrid method based on spectral kurtosis and ensemble empirical mode decomposition. Measurement 45:1308–1322
3. Yan R, Gao RX, Chen X (2014) Wavelets for fault diagnosis of rotary machines: a review with applications. Sig Process 96(5):1–15
4. Shen C, Wang D, Kong F et al (2014) Recognition of rolling bearing fault patterns and sizes based on two-layer support vector regression machines. Smart Struct Syst 13:453–471
5. Tamilselvan P, Wang P (2013) Failure diagnosis using deep belief learning based health state classification. Reliab Eng Syst Saf 115:124–135
6. Tran VT, Althobiani F, Ball A (2014) An approach to fault diagnosis of reciprocating compressor valves using Teager-Kaiser energy operator and deep belief networks. Expert Syst Appl 41:4113–4122
7. Gan M, Wang C, Zhu C (2016) Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. Mech Syst Signal Process 72–73:92–104
8. Qian N (1999) On the momentum term in gradient descent learning algorithms. Neural Netw Official J Int Neural Netw Soc 12:145–151
9. Nesterov Y (1983) A method for unconstrained convex minimization problem with the rate of convergence. In: Soviet Mathematics Doklady

An Automatic Feature Learning and Fault Diagnosis Method Based on Stacked Sparse Autoencoder

Yumei Qi, Changqing Shen, Jie Liu, Xuwei Li, Dading Li and Zhongkui Zhu

Abstract Fault feature extraction is the key for fault diagnosis, and automatic feature extraction has been a hot topic recently. Deep learning is a breakthrough in artificial intelligence and known as an automatic feature learning method. One of the most important aspects to measure the extracted features is sparsity. Thus, this paper presents a stacked sparse autoencoder (SAE)-based method for automatic feature extraction and fault diagnosis of rotating machinery. The penalty term of the SAE can help automatically extract sparse and representative features. Experiments and comparisons are conducted to validate the effectiveness and superiority of the proposed method. Results fully demonstrate that the stacked SAE-based diagnosis method can automatically extract more representative high-level features and perform better than the traditional intelligent fault diagnosis method like artificial neural network (ANN).

Keywords Feature extraction · Fault diagnosis · Sparse autoencoder Deep learning

1 Introduction

Rotating machinery is widely used mechanical equipment in the industrial field and of great importance to the economic development of the society. Unexpected machine faults may cause considerable economic losses even casualties. Therefore, numerous researchers have proposed various fault diagnosis methods to monitor the operation condition and improve the security and reliability of the rotating machinery [1]. Fault feature extraction is the key for fault diagnosis [2]. The quality of the extracted features directly determine the diagnosis accuracy. Wavelet transform, empirical mode decomposition, and statistical parameters analysis are

Y. Qi · C. Shen (✉) · J. Liu · X. Li · D. Li · Z. Zhu
School of Rail Transportation, Soochow University,
Suzhou 215131, Jiangsu, People's Republic of China
e-mail: cqshen@suda.edu.cn

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_39

the common used feature extraction methods. Although these methods have achieved a lot of success in feature extraction and fault diagnosis, they all rely on a large amount of manpower to extract discriminative features, which are time-consuming and require abundant expertise. Moreover, different features may lead to different diagnosis results, which is unstable [3].

Deep learning is a great breakthrough in machine learning, which can fully take advantage of its deep architectures to automatically extract essential high-level features. Compared to traditional feature extraction methods, deep learning methods are more time-saving and require no human interference. Up to now, deep learning has been successfully applied to speech recognition [4], computer vision [5], and natural language process [6]. These achievements fully illustrate the automatic feature learning ability of deep learning methods. Recently, deep learning is widely used for fault feature extraction and fault diagnosis. Chen et al. [7] employed a convolution neural network (CNN) for automatic fault feature extraction and fault diagnosis of gearboxes. Tamilselvan et al. [8] successfully diagnosed faults of aircraft and electric power transformer with features automatically extracted by a deep belief network (DBN). Jia et al. [9] successfully utilized a deep neural network (DNN) based on autoencoders (AE) for bearings and planetary gearboxes fault feature extraction and fault diagnosis. Shao et al. [10] put forward an AE-based method for feature extraction and fault detection. In addition to recognize the fault types, Guo et al. [11] proposed a CNN-based hierarchical fault diagnosis network for automatic bearing fault feature extraction and then diagnosed fault types and fault severities.

One of the most important aspects to measure the features automatically extracted by deep learning methods is whether these features can explore more useful information of the inputs and avoid redundancy to be representative, that is sparsity. Thus, a stacked SAE-based network is proposed in this paper for automatic fault feature extraction and fault diagnosis of rotating machinery. The penalty term of the SAE helps mine more abstract and representative features. Because bearings are vital elements of rotating machinery, which directly influence the operation condition of the rotating machinery, experiments and comparisons with bearing datasets are conducted to validate the effectiveness and superiority of the proposed method. The rest of the paper is organized as follows. Section 2 briefly introduces the stacked SAE network. Section 3 details the diagnosis flow chart of the proposed method. Experiments and comparisons are conducted in Sect. 4. Important conclusions are presented in Sect. 5.

2 A Brief Introduction to the Stacked SAE Network

2.1 SAE

An AE is an unsupervised feature learning neural network as depicted in Fig. 1. It consists of an encoder and a decoder network, where the former is responsible for learning features from the inputs automatically, whereas the latter is responsible for reconstructing the inputs from the learned features.

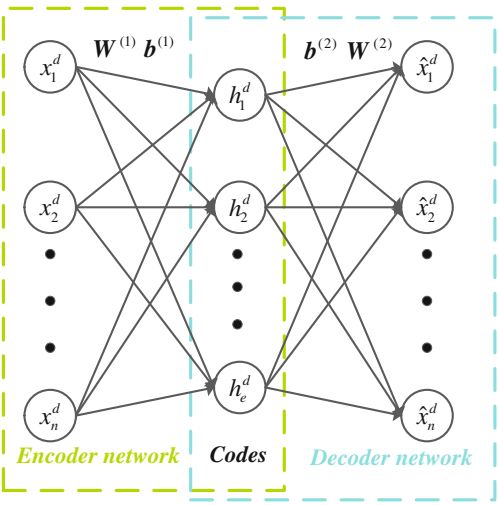


Fig. 1 Architecture of the AE

For each input vector $\mathbf{x}^d$ from datasets $\{\mathbf{x}^d\}_{d=1}^M$, the hidden vector $\mathbf{h}^d$ and the reconstructed vector $\hat{\mathbf{x}}^d$ can be defined as (1), where $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$ are the weight matrices, $\mathbf{b}^{(1)}$ and $\mathbf{b}^{(2)}$ are the bias vectors, and f is the active function. The overall cost function $J(\mathbf{W}, \mathbf{b})$ can be defined as (2), where λ is the weight decay parameter, n_l is the number of layers, S_l denotes the number of neurons in layer l, and $\mathbf{W}_{ji}^{(l)}$ is the connection weight between unit i and j in layer $l+1$ and l respectively.

$$\begin{aligned} \mathbf{h}^d &= f(\mathbf{W}^{(1)}\mathbf{x}^d + \mathbf{b}^{(1)}) \\ \hat{\mathbf{x}}^d &= f(\mathbf{W}^{(2)}\mathbf{h}^d + \mathbf{b}^{(2)}) \end{aligned} \tag{1}$$

$$J(\mathbf{W}, \mathbf{b}) = \left[\frac{1}{M}\sum_{d=1}^{M}\left(\frac{1}{2}\left\|\mathbf{x}^d - \hat{\mathbf{x}}^d\right\|^2\right)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{S_l}\sum_{j=1}^{S_{l+1}}\left(\mathbf{W}_{ji}^{(l)}\right)^2 \tag{2}$$

A SAE can be obtained by applying a sparsity penalty term to the cost function $J(\mathbf{W}, \mathbf{b})$ of the AE, and the overall cost function of the SAE is defined as (3), where ρ is the sparsity parameter, $\hat{\rho}_g$ is the average output of the hidden unit g, and β is the sparsity penalty parameter.

$$J_{sparse}(\mathbf{W}, \mathbf{b}) = J(\mathbf{W}, \mathbf{b}) + \beta\sum_{g=1}^{S_2}\left(\rho\log\frac{\rho}{\hat{\rho}_g} + (1-\rho)log\frac{1-\rho}{1-\hat{\rho}_g}\right) \tag{3}$$

The training process of the SAE aims to automatically learn more sparse and representative features of the original inputs. After the parameters initialization, back propagation (BP) algorithm is used to minimize the overall cost function of the

SAE and update the parameters until reaching the maximum iteration. The optimal parameter sets $\mathbf{W}^{(1)}$, $\mathbf{W}^{(2)}$, $\mathbf{b}^{(1)}$, and $\mathbf{b}^{(2)}$ can be learned simultaneously during the training process and a well-trained SAE can be obtained finally.

2.2 Stacked SAE Network

U SAEs can be stacked to construct a stacked SAE network with U hidden layers, as shown in Fig. 2. The softmax classifier is added to the output layer for classification.

The training process of the stacked SAE network consists of two main procedures, namely, pre-training and fine-tuning. The pre-training aims to pre-train each SAE and the softmax classifier at a time to learn the intra-relationship of each model and obtain a set of optimal weight initialization, whereas the fine-tuning aims to learn the inter-relationship among layers and establish the most suitable relationship between the learned features and the labels. First, BP algorithm is used to pre-train SAE1, SAE2, and until the Uth SAE is well trained. Then the softmax classifier can be well trained through supervised learning. Finally, BP algorithm is used to fine tune the whole network to establish the most suitable relationship between the learned features and the labels.

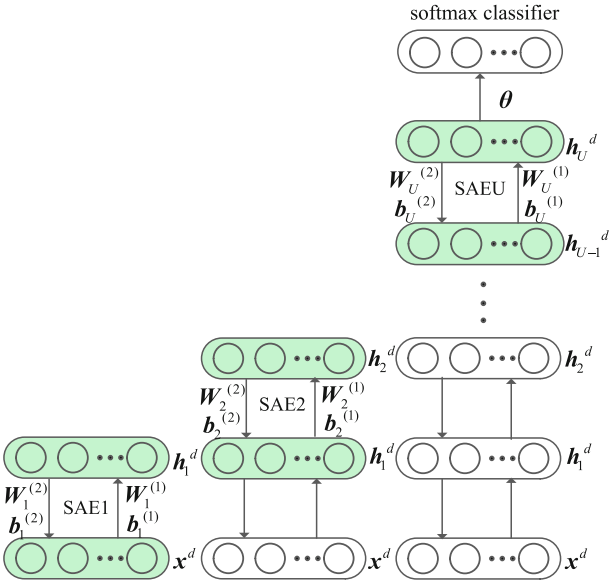


Fig. 2 Architecture of the stacked SAE network

3 Proposed Stacked SAE-Based Fault Diagnosis Method

The flow chart of the proposed stacked SAE-based diagnosis method is shown in Fig. 3. The collected signals are firstly transformed to the frequency domain by the Fast Fourier Transform (FFT). Then the datasets are randomly divided into training and testing two datasets, where the former and the latter are respectively responsible for training and validating the constructed stacked SAE-based diagnosis network.

4 Experiments for Validation

4.1 Data Introduction

Bearing fault diagnosis is analyzed to demonstrate the effectiveness of the proposed method. Figure 4 shows the test rig. The defect was artificially set at surfaces of the inner race, outer race, and the rolling element with a width of 0.20 mm. Four health condition datasets, namely, inner race fault (IF), outer race fault (OF), ball fault (BF) and normal condition (N) were obtained with a sampling frequency of 10 kHz and a data sampling point of 1024, which are labeled as 1, 2, 3, 4, respectively.

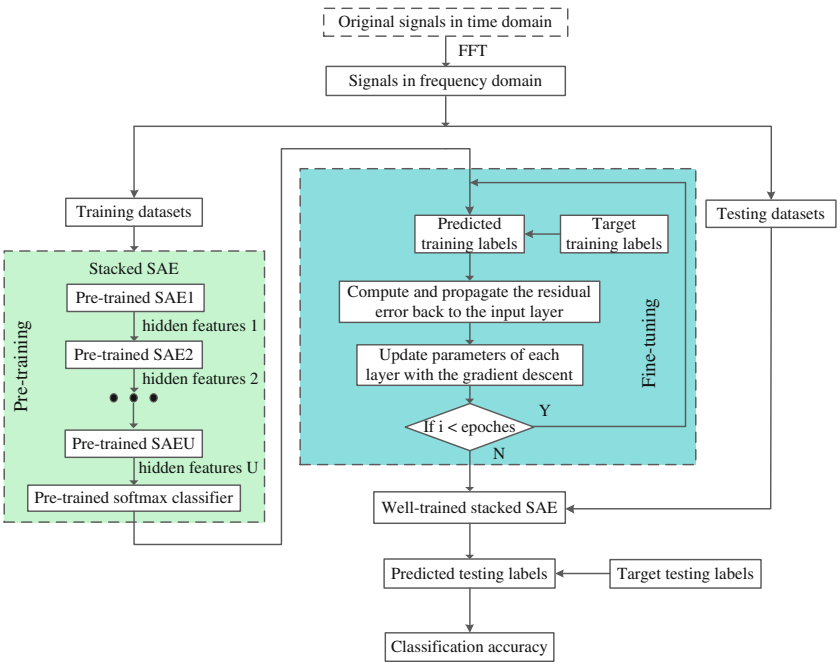


Fig. 3 Diagnosis flow chart of the proposed stacked SAE-based diagnosis network

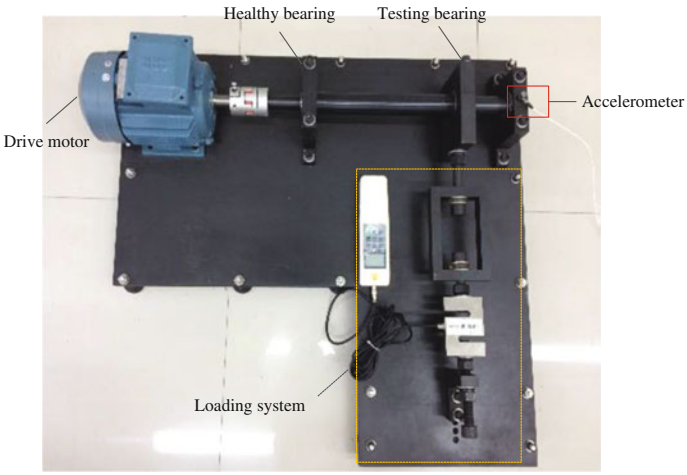


Fig. 4 Test rig for bearing

Each health condition datasets contained 600 samples, where 300 samples randomly selected and the remaining 300 samples are regarded as the training and testing datasets.

4.2 Validation Results

FFT is used to transform the collected bearing signals to the frequency domain, and the data points in frequency domain are chosen to be 1024. Because the first 512 points of the amplitude spectrums are completely symmetrical with the last 512 points, only the first 512 points are regarded as the inputs of the proposed diagnosis network. Thus, the input size of the network is 512. Table 1 shows the detailed parameters of the constructed stacked SAE network. The iteration numbers of each SAE, softmax classifier and the fine-tuning process is set as 8000, 8000, 8000, 10,000, and 10,000, respectively. The proposed method achieves an amazing testing accuracy of 100%.

To further demonstrate the superiority of the proposed method in automatic representative feature extraction and fault diagnosis, stacked AE network with the same architecture of the stacked SAE network and the ANN are applied for

Table 1 Detailed parameters of the proposed diagnosis network

No. of hidden layers	Hidden layer size	Input layer size
3	200-100-50	512
Output layer size	Learning rate of each layer	Learning rate of fine-tuning
4	0.3-0.3-0.3-0.2	0.2

comparison. The ANN has only one hidden layer, which consists of 50 neurons. The detailed testing accuracy is shown in Table 2. The first three methods are applied with the amplitude spectrums, whereas the rest methods deal with the original time-domain signals directly. By observation, the stacked SAE and the AE diagnosis network perform better than the ANN using the same inputs, no matter the inputs are in time domain or frequency domain. Moreover, the superiority is significant especially when handling the original time-domain signals, which fully validates the superiority of deep learning methods in automatically extracting more discriminative high-level features useful for classification. In addition, the stacked SAE-based diagnosis network performs better than the stacked AE-based diagnosis network using the same inputs, which fully illustrates the superiority of the SAE in mining more sparse and representative features due to its sparsity penalty term than the AE. Meanwhile, methods dealing with amplitude spectrums perform better than the same methods dealing with the original time-domain signals, which demonstrates that the amplitude spectrums are good low-level features and able to help the network further extract more representative and useful features. In general, the proposed stacked SAE-based diagnosis method dealing with the amplitude spectrums performs the best.

In addition, principle component analysis is applied to visualize the extracted features as shown in Fig. 5. In Fig. 5b, c, the features mined by the deep learning networks that represent the same health condition are gathered better, and those

Table 2 Testing accuracy of the comparative methods

Methods	IF	OF	BF	N	Total
FFT+SAE	100	100	100	100	100
FFT+AE	100	100	99.67	100	99.92
FFT+ANN	100	100	99.33	100	99.83
SAE	68.33	36.33	54.00	66.67	56.33
AE	46.67	70.00	45.67	56.00	54.58
ANN	19.33	28.67	33.67	40.00	30.42

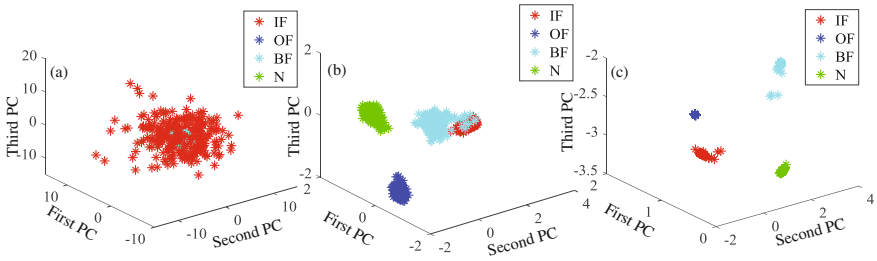


Fig. 5 Features visualization: **a** features in the time domain, **b** features mined by the stacked AE network, **c** features mined by the stacked SAE network

represent the different health conditions are separated better compared to the plot in Fig. 5a, which fully demonstrate the automatic feature learning ability of deep learning methods. In Fig. 5b, some features that represent the IF and BF are mixed, whereas in Fig. 5c, all the features representing the four health conditions are separated well, which fully demonstrates the superiority of SAE in extracting more sparse and representative features than the AE. Generally, the stacked SAE-based diagnosis network can automatically extract more sparse and representative features, which can remain as much useful information of the inputs as possible and is effective enough for fault classification.

5 Conclusions

This paper proposes a method based on stacked SAE network for automatic sparse and representative feature extraction and fault diagnosis of rotating machinery. The inputs of the diagnosis network are the amplitude spectrums. Experimental analysis and comparisons using bearing datasets are carried out to validate the effectiveness of the proposed method. Results fully show the automatic feature learning ability of deep learning methods, compared to the ANN with shallow architectures. Besides, it proves the superiority of the SAE in extracting more sparse and representative features, compared to the AE. Meanwhile, the amplitude spectrums are proved to be good low-level features, which are able to help the network further extract more representative and useful features than those extracted from the time-domain signals.

Acknowledgements This research was financially supported by the National Natural Science Foundation of China (Grant No. 51505311 and 51375322), the Natural Science Foundation of Jiangsu Province (No. BK20150339) and the China Postdoctoral Science Foundation funded project (2016T90490).

References

1. He Q, Wang J, Liu Y et al (2012) Multiscale noise tuning of stochastic resonance for enhanced fault diagnosis in rotating machines. Mech Syst Signal Process 28:443–457
2. He Q, Ding X (2016) Sparse representation based on local time–frequency template matching for bearing transient fault feature extraction. J Sound Vib 370:424–443
3. Lu C, Wang Z, Zhou B (2017) Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. Adv Eng Inform 32:139–151
4. Hannun A, Case C, Casper J et al (2014) Deep speech: scaling up end-to-end speech recognition. Computer Science
5. Toshev A, Szegedy C (2014) Deeppose: Human pose estimation via deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1653–1660

6. Johnson R, Zhang T (2014) Effective use of word order for text categorization with convolutional neural networks. Eprint Arxiv
7. Chen ZQ, Li C, Sanchez RV (2015) Gearbox fault identification and classification with convolutional neural networks. Shock Vib 2:1–10
8. Tamilselvan P, Wang P (2013) Failure diagnosis using deep belief learning based health state classification. Reliab Eng Syst Saf 115:124–135
9. Jia F, Lei Y, Lin J et al (2016) Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech Syst Signal Process 72:303–315
10. Shao H, Jiang H, Zhao H et al (2017) A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. Mech Syst Signal Process 95:187–204
11. Guo X, Chen L, Shen C (2016) Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement 93:490–502

Application of Intelligent Hanging Production System in Garment Industry

Jishu Zhang, Yingzi Zhang and Peiguo Wang

Abstract The intelligent clothing hanging production system is an efficient new fashion production method to meet the current production practice. The development history and production structure of intelligent clothing suspension system are summarized. The process adaptability of the system and the requirement of process balance in the application process are discussed. The results showed that the clothing hanging production system has basically realize intelligence after many years of development. This system is an elastic combination system composed of transmission hardware and control software, and through the coordination of master control software and transmission hardware, it can be achieved easily that real-time management, automatic balance material, automatic material sorting, parallel operation of multiple styles process. Compared with the traditional garment production methods, the effective use of intelligent clothing hanging production system can improve the production efficiency of at least 20–30%.

Keywords Intelligent hanging production system · Clothing intelligent manufacturing · Lean production

1 Introduction

The types and styles of garments are developed to the trends of increasing variety, and the orders of factories are also showing the trend of small batch production [1]. The accelerated fashion cycle has forced the production cycle to be shortened. Therefore, the traditional mode of production in the clothing industry has been seriously challenged, and has not been able to meet the actual needs of today's production. It has become an important project for people to actively explore new ways of garment processing and production [2]. In such a situation, the intelligent clothing hanging production system is developed [3]. Compared with the traditional

J. Zhang (✉) · Y. Zhang · P. Wang
School of Art and Fashion Engineering, Changshu Institute of Technology, Changshu, China
e-mail: zhangjishu@163.com

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_40

"banding" mode of production, clothing hanging transmission system can generally improve the production efficiency of 20–30% [4]. The system is well suited to the market demand of small volume, variety, short cycle and quick response in the contemporary clothing market.

In order to fully display the advantages of intelligent clothing hanging production system, the development of intelligent garment hanging system and its production structure are presented, the adaptability of the system and the requirements for the balance of the process in the application process are discussed. It will provide reference for promoting the development of the system and application of garment industry.

2 Development Process and Working Principle of Intelligent Clothing Hanging Production System

2.1 Development Process of Intelligent Clothing Hanging Production System

The intelligent clothing hanging production system is a kind of rapid reaction production technology used in the field of garment production. It is the concentrated expression of mechatronics technology in garment production. In 1970s, the Swedish company ETON shirts factory has developed the first hanging transportation production system [5]. Through continuous improvement and innovation, the second generation ETON2001 transmission system and the third generation ETON2002 transmission system were launched in 1983 and 1987, respectively, which are the rudiment of modern intelligent suspension system. After ETON Co. Ltd., the Japanese BROTHER Co. Ltd., JUKI Co. Ltd., GERBER Co. Ltd. of USA, DUERKOPP Co. Ltd. of Germany, LECTRA Co. Ltd. of France and Spain INVESRTONIAC Co. Ltd. also has developed a suspension transmission system of their own, and constantly improve the mechanical transmission parts. In China, the research on the clothing hanging production system is also moving forward [6–8]. For instants, in 2012, Jiangsu Nantong Ming Xing Technology Development Co. Ltd. and Chinese Academy of Sciences successfully applied RFID recognition technology to the production line management in the field of clothing hanging transmission, which solved the "Bottleneck" problem of the production process of knitting and woven garment industry. The clothing hanging production system is moving from intelligent one to smart one.

2.2 Working Principle of Intelligent Clothing Hanging Production System

The clothing hanging production management system has two parts, hardware and software, and the hardware part is connected by each workstation (station) arranged

in accordance with the process requirements [9]. It can connect different processes, different levels, and even workstations between different buildings, so that products can be easily transmitted. The software part is composed of a central controller in the form of a network connecting intelligent terminal every workstation and through the central master computer screen can immediately see the current production condition, so as to achieve the control of the whole production line, which make the production management have relatively good visibility.

According to the requirements of the processing technology in production, all the pieces of the whole garment are hung on the clothes hanger, arranged according to the prior arrangement of the man-machine working station. The piece automatically transported to the operator of next procedure through the intelligent transport system, and directly delivery to the most convenient location for each operator. When the production staffs complete the process, only the light control button is pressed, and the hanging system can automatically transfer the hanger with the semi-finished garment to the next workstation. It minimizes the non-production utilization time of handling, bundling and folding. The system is a kind of elastic combination system, which is not limited by the length of processing lines and the arrangement of processing positions and can be long, short, high or low. It can be arranged and assembled autonomously according to the current needs, and it is flexible and convenient. The system is especially suitable for various production occasions requiring continuous processing and automation of transportation process.

It is not difficult to see that the whole sewing process of the intelligent hanging system of the garment is carried out in the state of suspension, and the processing process needs no binding, handling, unpacking and other floating actions. In this way, the problems of traditional production mode, such as more moving time, more occupancy, long production cycle and difficult quality control, are solved effectively.

3 Superiority Analysis of Intelligent Clothing Hanging Production System

3.1 Parallel Production into Multiple Styles

The production line requires simultaneous production of two or more styles when the product process is relatively small or the order quantity suddenly increases, and it also ensures lower machine idle rate and no increase in production line. At this point, the clothing hanging production system can solve this problem more easily. It can divide the workstation of the hanging line into two groups or several groups through the control system, and control the transmission route of the rack on the transmission frame by two or more process control procedures. The different lines of production are in a relatively independent state.

When the production process and styles of clothing are very similar, even if the two products are different, it can also be achieved to put them into the same production line simultaneously based on the production practice. Under the control of hanging system, different styles of products are transmitted to specific station from different transport lines to the operator for continuous processing of different styles of material. After the operator of the station carries on the continuous processing to the different style material, these garments are transmitted to the different inspection table to carries on the classification inspection.

3.2 Computer Real-Time Management

Of course, the problems encountered above are not only produced at the same time, but also encountered in the production process at ordinary times. This is called "bottleneck problem". The administrator of main control computer of the hanging production system, through the master computer, can understand the real-time information produced by each station, and carry on the existing pipeline adjustment to achieve real-time balanced production.

At the same time, through the system, we can also understand the current production model and each process production output status in real time, so as to make real-time adjustments. In addition, the computer host can also check the workload of the staff, so production can be easily learned. As for the QC, when the inspectors find that the garment is defective, and need to rework, QC personnel only need to hand control box button of Rework, the system will automatically transfer the defective garment to workstation location to rework, greatly improving the rework efficiency.

The computer mainframe of the hanging system is usually networked with the whole factory computer management system, so as to realize the advanced management inside the whole enterprise, which is the indispensable equipment for the garment enterprise to realize the information manufacture.

3.3 Material Automatic Sorting

According to the actual situation, the intelligent garment hanging production system can sort production material category automatically according to grade level. And it also can distinguish materials with different colors and sizes. Therefore, in this way, the system can set a special process for each production with different type or color. Each cut-piece of different varieties or the color in this production line will have their individual process flow which protect the running will not happen chaos collision effectively.

3.4 Automatic Balancing Distribution of Materials

Due to the different technical and ability of the factory employees, and the types of materials needed to be processed are different, the intelligent hanging production system will automatically adjust the number of the racks in the work station. If the number of pieces in the workstation has reached full load conditions, the system will automatically transfer piece hangers needing to be processed into the buffer area of the hanging system, until the workstation is not in full or after normal water flow, the system then automatically transports the cut-piece in the buffer to the workstation, which can effectively prevent pipeline plugging

4 Production Efficiency Analysis of Garment Intelligent Suspension Production System

As an example, 50 shirts were produced in two ways: Traditional production method and Hanging production method. The production take time was 0.7 min, and there were 7 production stations in line. According to the traditional method of production, a basket of materials carried to the next process after a process done until the product is finished. Without considering loading, handling, and other ancillary operations time, its production cycle is $T1 = 0.7 \times 50 \times 7 = 245$ min. In the same manner as the traditional mode of production, If using garment hanging production system, its production cycle is $T2 = 0.7 \times 7 + (50-1) \times 0.7 = 39.2$ min, which is shorten more than using traditional method of production. That is because parallel production is achieved due to the parallel movement of the hanging production line.

Not only that, the hanging system in material transportation compared with the traditional production methods has a substantial reduction in the auxiliary time, only this point it can make production efficiency increase by about 20%.

Through the actual production observation and statistical data analysis, the production process time assignment is obtained, as is shown in Table 1.

From Table 1, it can be seen that the efficiency of the basic operation of hanging production system is more than the traditional production modes, and allowance rate of the former is much lower than the production efficiency of the later. As we

Table 1 Comparison of efficiency of intelligent clothing suspension system and traditional production methods

Form of production line	Efficiency of intelligent clothing suspension system (%)	Efficiency of traditional production methods (%)
Basic operation	38.9	26.3
Auxiliary operation	45.8	42.3
Allowance rate	15.3	31.4

all know, the production efficiency of factory is positively proportional to the basic operation rate, and is inversely proportional to the allowance rate. From the above two points, we can see that the auxiliary intelligent hanging line has absolute advantages for improving production efficiency.

5 Conclusion

In summary, the intelligent garment hanging production system has its unique performance, which can control production by the main control computer, achieve simultaneous implementation of processes and real-time control of the production, adjust the production line in time, and can transform processing varieties and the number of production lines. The intelligent garment hanging production system can adapt well to the present production status of clothing: fast-paced, multi style and small batch mode, and is an innovative technology application for the apparel industry. The practice has showed that the use of intelligent garment hanging production system has greatly improved the production efficiency of the enterprise.

References

1. Chen XM (2015) Analysis of present situation and innovation development of China's clothing industry. J Managers 15:187–188
2. Chen SB (2012) Automatic transmission system in the production of clothing in the applied research. J Manufact Autom 34(4):151–153
3. Yuying Qiu (2008) Application of FMS to product assembly line. J Jiaxing Univ 26(2):123–125
4. Chen JY, Chen RK, Xu YY (2009) Intelligent clothing hanging system to improve efficiency. Text Apparel weekly, 72
5. Cai WJ, Jing Jin (2015) General situation and application research of clothing suspension transmission system. Light Ind Sci Technol 208(8):115–116
6. Xia Chen, Sun WS (2014) Study and practice of load balancing design of knitwear in overhead sewing system. Knitting Ind 8:55–59
7. Chen HQ, Chen Y, Ding J et al (2012) Analysis of garment hanging production line organization. Silk 49(01):30–32
8. Liu AL, Lin ZH, Sun YZ (2017) Sliding mode control in clothing hanging transportation system based on load observer. Wool Text J 36(4):56–61
9. Lei M (2017) Application of computerized hanger system in textile industry. China Text Leader 36(4):58–60

Objective Evaluation of Clothing Modeling Based on Eye Movement Experiment

Yingzi Zhang and Jishu Zhang

Abstract The evaluation of clothing beauty has always been the focus of attention of designers and consumers, and subjective description is the main method of evaluation, but lack of effective objective evaluation method. In this paper, 15 different tests and 4 different fashion pictures were selected, and the eye movement data of different testers on garment modeling were collected and analyzed by means of Eye tracker, and the data was compared with the subjective evaluation in the traditional context. The results showed that the result of eye movement data analysis is consistent with the language description, which shows that the method of garment modeling evaluation based on eye movement experiment is a feasible objective evaluation method.

Keywords Eye movement experiment · Evaluation of fashion beauty
Kansei engineering

1 Introduction

The comprehensive evaluation of garment modeling beauty is formed by direct perception of the detail elements of costume modeling through the human senses (sight, touch), and is described in terms of language and text coming from the customer's judgment of the overall beauty of garment styling. These evaluations are usually based on subjective qualitative descriptions, such as coordination or abrupt in color matching, fit or lack of fit in pattern, formality or exaggeration in profile, fine or messy designs in design, and so on, which is inaccurate and unquantitative [1]. When describing the difference of clothing beauty, there may be situations such as subjective cognition and inconsistency of expression language, and inability to express in words occasionally. With the rise of Kansei engineering concept, fashion designers began to use the term of engineering expression method to quantify the

Y. Zhang (✉) · J. Zhang
School of Art and Fashion Engineering, Changshu Institute of Technology, Changshu, China
e-mail: zhangyingzi92@163.com

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_41

rational description of previous sensory evaluation and expression, and try to grasp consumers through quantitative sensory, so as to design the product that consumers love [2–4].

Eye movement experiment is an experimental method commonly used in Kansei engineering research [5–7]. Through analyzing and recording human eye movement data, the relationship between eye movement and human psychological activity is explored, and the consumer psychology is grasped. If it can be used in the evaluation of costume beauty to form a set of objective evaluation of costume beauty method, not only can expand the way of garment design evaluation, but also expand to other clothing related application fields. In this paper, different testers of the costumes and effect of the eye movement data were collected and analyzed by the eye movement experiment, and the data was compared with the traditional subjective evaluation context, and tries to construct the objective evaluation method of costume beauty to provide objective reference for fashion design.

2 Experiments

2.1 Material Preparation

Select a group of design works with "time" as the theme of design, which composed of four styles of design renderings. For ease of description, clothing numbers are 1, 2, 3, 4 from left to right (Fig. 1).

2.2 Instrument Preparation

German SMI desktop eye tracker was selected. Working distance was 50–80 cm, high accuracy was 0.5°, spatial resolution (RMS) was 0.1°, sampling rate was 60 and 120 Hz, the time of fast automatic calibration was <3 s (2 points calibration).

Fig. 1 Four design works with "TIME" as the theme of design

2.3 Experiment Scheme

A single factor test method was used to test the experimental samples to ensure the consistency of the contrast pictures of the group. In the experiment process, according to the order of design start from one to another, and the pictures appear randomly within the given time. According to the experimental requirements, the pictures were prearranged, and the five design factors are compared including the collar, shoulder, sleeve, pattern and contour.

2.4 Test Methods and Requirements

(1) The test requires students from freshman to senior, who have major in clothing and non clothing. Sample size is 15.
(2) Normal or corrected to normal vision, need to wear glasses please must bring glasses (please bring glasses, do not take concealed glasses).
(3) Asked the subjects to sit in a relatively comfortable position with eyes up to the center of the screen, 60–80 cm away from the screen, and try not to move as soon as they were seated.
(4) Put four groups of pictures into the computer, and the pictures are rendered at intervals of 30 s. Inform the testers of the test contents before each test is initiated.
(5) At the end of each test, the tester is asked to answer the test questions in language.
(6) In order to facilitate the quantitative description of effect evaluation, four attention levels are set 0, 1, 2, 3, 0 of them show no attention, 1 show general concern, 2 show more attention, and 3 show the most concern.

3 Results and Discussions

3.1 Results and Analysis of Eye Movement Test

3.1.1 Analysis of the Degree of Attention of Each Garment Styling Element

Table 1 showed the results of attention measurement of garment modeling elements based on eye movement experiment data. As can be seen from the above table, in the collar area attention degree, the design effect diagrams 2 focuses on the strongest, and secondly is the effective drawing 1. That may be about color matching. Effective drawing 1 is dominated by red color, with less obvious changes in adjacent colors, or near the color purity of the colors, which slowly retain the

Table 1 Experimental results of attention to garment modeling elements based on eye movement experimental data

Styling elements of concern	Number of garment effect drawing	Degree of concern
Collar	1	0 1 2√ 3
	2	0 1 2 3√
	3	0 1√ 2 3
	4	0 1√ 2 3
Shoulder	1	0 1 2 3√
	2	0 1√ 2 3
	3	0 1√ 2 3
	4	0√ 1 2 3
Sleeve	1	0 1 2√ 3
	2	0 1√ 2 3
	3	0√ 1 2 3
	4	0 1√ 2 3
Pattern	1	0 1 2√ 3
	2	0 1 2 3√
	3	0 1√ 2 3
	4	0 1√ 2 3
Contour	1	0 1 2 3√
	2	0 1 2√ 3
	3	0 1√ 2 3
	4	0 1 2√ 3

attention of the subject. In the shoulder area attention, effective drawing 1 is the strongest concern, followed by effective drawing 2. In the attention of the sleeve, the attention in the four renderings is not high, less attention. In profile attention, the effects of effective drawing 1 and effective drawing 2 are the most important. Finally, on the attention level of the pattern, effective drawing 1 is the most concerned among all the concerned areas, followed by effective drawing 2 and effective drawing 4.

3.1.2 Analysis of Overall Attention Degree of Clothing Modeling Elements

Table 2 presents the overall attention of clothing modeling elements. As can be seen from the table, among the 5 elements of the selection of design, the attention of the elements of the pattern is 8, which gets the most attention, followed by profile and collar, and the least concern is the sleeve. The overall attention degree of garment modeling elements shows the importance of each factor in garment modeling. It is not difficult to see that contour occupy an important position in the fashion design of this group.

Table 2 Overall attention degree of clothing modeling elements

Styling elements of concern	Collar	Shoulder	Sleeve	Profile	Contour
The sum of concerns	7	5	4	7	8

3.1.3 Analysis of Overall Attention Degree of Clothing Modeling Elements

Table 3 shows the overall focus of each garment. As can be seen from the table, among the four effective drawings, the attention of picture 1 is 12, which got the most attention, followed by picture 2. And the minimum degree of attention is picture 3. The collar, shoulders, sleeves, profile and patterns have different effect on the clothing design, thus affecting the overall evaluation of the costumes on the beauty of the clothing. The general attention is on the test item of clothing modeling elements and the degree of concern, which maps out the comprehensive evaluation of the costumes. The higher the overall attention of clothing is, the more acceptable to the design the testers are. Therefore, in order to achieve the best effect of clothing design, designers should not only pay attention to many design elements, but also handle the relationship between other elements, both primary and secondary, and make them in a stable and harmonious state.

3.2 Results and Analysis of Language Description of Testers

3.2.1 Subjective Description of the Degree of Concern in Modeling Elements

Table 4 presents the experimental results of modeling attention based on the subjective description of testers. Comparing Table 4 and Table 1, it is not difficult to see that except for few formative factors, the subjective descriptions of the other modeling factors and the results of eye movement experiments have high similarity. According to statistics, the similarity reached 94%.

3.2.2 Subjective Description of the Overall Attention Level of the Design Renderings

Table 5 gives the number of people who approve of the effective drawing. It can be seen from Table 5, picture 1 was expressed approval by most testers subjectively. Most testers think that the use of the contour design in the neckline easy to focus

Table 3 Overall focus of each garment

Number of garment effect drawing	1	2	3	4
The sum of concerns	12	10	4	5

Table 4 The results of model design focus based on the subjective description of the testers

Styling elements of concern	Number of garment effect drawing	Degree of concern
Collar	1	0 1 2√ 3
	2	0 1 2 3√
	3	0 1√ 2 3
	4	0 1√ 2 3
Shoulder	1	0 1 2 3√
	2	0 1√ 2 3
	3	0 1√ 2 3
	4	0√ 1 2 3
Sleeve	1	0 1 2√ 3
	2	0 1√ 2 3
	3	0√ 1 2 3
	4	0 1√ 2 3
Pattern	1	0 1 2 3√
	2	0 1 2 √3
	3	0 1√ 2 3
	4	0 1√ 2 3
Contour	1	0 1 2 3√
	2	0 1 2√ 3
	3	0 1 2√ 3
	4	0 1√ 2 3

Table 5 Statistics on the number of people who approve of the effect

Number of garment effect drawing	1	2	3	4
The number of people who approve of the effect	6	4	2	3

sight, the design of profile simple, the proportion of color design harmonious, asymmetric shoulder sleeve design adding dynamic. But style 3 was approved by the minimum testers, in the eye of most people, it is concurred that the design is more cumbersome and has a weight imbalance.

4 Conclusions

By comparing the experimental results of eye movement experiments and subjective descriptions, we can draw the following conclusions:

1. The similarity between the eye movement experiment and the subjective description is near 94%. It can be considered that the testers have a high degree

of consistency in the attention of the four modeling elements,Which showed that the objective evaluation of garment modeling elements based on eye movement experiment is feasible.
2. The results of eye movement experiment objectively show the weight evaluation data of the elements of garment modeling design.
3. The eye movement experiment does not need the tester to carry on the language description, which objectively shows the evaluation of the overall beauty of clothing modeling.

5 Compliance with Ethics Standards

1. Disclosure of potential conflicts of interest: The authors declare that they have no conflict of interest.
2. Research involving human participants: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.
3. Informed consent: Informed consent was obtained from all individual participants included in the study.

References

1. Ye LC (2011) Athletics of customs. China Textile & Apparel Press, Beijing
2. Chen YJ, Wang PG (2013) Discussion on detail design of knitted apparel based on eye tracker. J Mod Text Technol 6:28–31
3. Dan Shao, Zhu LS (2013) Research on image cognitive of branded apparel based on eye-tracking technology—case study of E brand product style analysis. J Donghua Univ Nat Sci 39(2):240–246
4. Li CH, Yan L (2012) Kansei Engineering and its application in clothing style evaluation. Tianjin Text Sci Technol 128:28–30
5. Zheng JJ, Ji XF (2016) Visual perception of costumers on clothing display. J Text Res 37 (3):161–165
6. Wang YZ, Cai QY, Zhang JN et al (2016) Recognizing costume features of She nationality in different areas based on eye-moving experiment. Silk 53(6):32–37
7. Wang Y, Lv FQ (2017) Usability testing of ATM machines interface based on eye-tracking data. Chinese J Ergon 32(1):48–54

Simulation Analysis of Rotor Indirect Field Oriented Vector Control System for AC Induction Motor in Low Speed Electric Vehicles

Kang Wang, Qingzhang Chen and Zhengyi Wang

Abstract This paper establishes the mathematical model of the AC induction motor, and according to the basic principle of rotor indirect field oriented vector control, a simulation model of the AC induction motor control system in low speed electric vehicles was established with the help of MATLAB/Simulink software. Simulation results show that this control used in a low speed electric vehicle has excellent static stability and dynamic response, such as the speed and torque response is quick, and the speed overshoot is small, which reveals the practicality and validity of the developed model, the results also provides a foundation for design and debug of low speed electric vehicles motor control system.

Keywords Low speed electric vehicles · The AC induction motors
Vector control · MATLAB simulation

1 Introduction

Electric vehicles motor should with a wide range of speed, high efficiency, large instantaneous power, strong overload capacity, better environmental adaptability and other performance requirements. And because of the AC induction motor with the advantages of low price, small size, light weight, reliable operation, better environmental adaptability, small torque ripple, it is used in electric motor drive system very early.

K. Wang
College of Mechanical and Electrical Engineering, Soochow University,
Suzhou 215000, China
e-mail: autowk@126.com

Q. Chen (✉) · Z. Wang
School of Automotive Engineering, Changshu Institution of Technology,
Suzhou 215500, China
e-mail: cqz@cslg.edu.cn

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_42

The vector control realizes the decoupling control of the flux and torque of AC induction motor, and it's speed control system also can provide the dynamic performance required for the electric vehicles [1]. These methods can be mainly divided into two categories, one is direct magnetic field directional control, which needs to directly measure the rotor magnetic field and also easily affected by other parameters of the motor. And the other is indirect magnetic field directional control, which doesn't need to directly measure the rotor magnetic field but controls the rotate difference frequency control to achieve directional control, and this control method is easier to implement in execution. However, taking into account the control system is more complex, as well as reduce R & D costs, shorten the R & D cycle, improve control performance and other factors, we use computer technology and simulation technology to control the system modeling and simulation analysis.

Therefore, this paper establishes the AC induction motor and the mathematical model of direction vector control based on rotor indirect magnetic field. The MATLAB/Simulink software is used to establish the simulation model of the AC induction motor control system for the low speed electric vehicles, and the simulation results are analyzed.

2 Mathematical Model of the AC Induction Motor

Make the AC induction motor rotor equal to three-phase wound rotor, then convert it to the stator side, the stator and rotor of after conversion turns are equal. And make the following assumptions [2]:

(a) Ignoring spatial harmonics. The three-phase stator windings (ABC) and the three-phase rotor windings (abc) differ by 120° in the space, and the resulting magnetic potential is distributed along the air gap in the light of sine distribution.
(b) Ignoring magnetic saturation.
(c) Ignoring core loss.
(d) Ignoring the influence of frequency and temperature on winding resistance.

Physical model of three-phase AC induction motor in Fig. 1. The stator three-phase winding axis A, B and C are fixed in the space, the A axis is the reference axis, and the rotor three-phase winding axis a, b and c rotate with the rotor, and the angle between the rotor a axis and the stator A axis is the space angular displacement variable θ. The voltage equation, the flux linkage equation, the torque equation and the motion equation together form the mathematical model of the AC induction motor.

(a) The voltage equation. According to the voltage balance equations of three-phase stator winding and the voltage equation of three-phase rotor winding converted to stator side, the voltage equation of AC induction motor can be obtained:

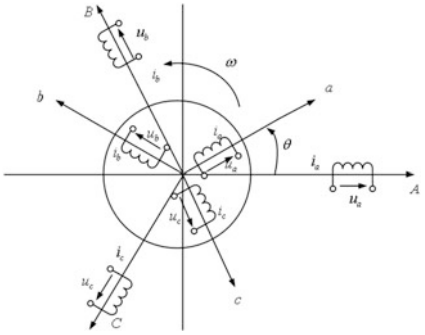


Fig. 1 Physical model of three-phase AC induction motor

$$\begin{bmatrix} u_A \\ u_B \\ u_C \\ u_a \\ u_b \\ u_c \end{bmatrix} = \begin{bmatrix} R_s & 0 & 0 & 0 & 0 & 0 \\ 0 & R_s & 0 & 0 & 0 & 0 \\ 0 & 0 & R_s & 0 & 0 & 0 \\ 0 & 0 & 0 & R_r & 0 & 0 \\ 0 & 0 & 0 & 0 & R_r & 0 \\ 0 & 0 & 0 & 0 & 0 & R_r \end{bmatrix} = \begin{bmatrix} i_A \\ i_B \\ i_C \\ i_a \\ i_b \\ i_c \end{bmatrix} + p \begin{bmatrix} \psi_A \\ \psi_B \\ \psi_C \\ \psi_a \\ \psi_b \\ \psi_c \end{bmatrix} \tag{1}$$

In Eq. (1):

$u_A, u_B, u_C, u_a, u_b, u_c$ represent the instantaneous values of the stator and rotor phase voltages, respectively.
$i_A, i_B, i_C, i_a, i_b, i_c$ represent the instantaneous values of the stator and rotor phase currents, respectively.
R_s, R_r represent stator and rotor windings, respectively.
p represents the differential operator, instead of $\frac{d}{dt}$.

Equation (1) simplified:

$$u = Ri + p\psi \tag{2}$$

(b) The flux linkage equation. The AC induction motor for each winding flux linkage is the sum of the self-inductance flux linkage and other winding pairs of AC induction motor mutual inductance flux linkage.

$$\begin{bmatrix} \psi_A \\ \psi_B \\ \psi_C \\ \psi_a \\ \psi_b \\ \psi_c \end{bmatrix} = \begin{bmatrix} L_{AA} & L_{AB} & L_{AC} & L_{Aa} & L_{Ab} & L_{Ac} \\ L_{BA} & L_{BB} & L_{BC} & L_{Ba} & L_{Bb} & L_{Bc} \\ L_{CA} & L_{CB} & L_{CC} & L_{Ca} & L_{Cb} & L_{Cc} \\ L_{aA} & L_{aB} & L_{aC} & L_{aa} & L_{ab} & L_{ac} \\ L_{bA} & L_{bB} & L_{bC} & L_{ba} & L_{bb} & L_{bc} \\ L_{cA} & L_{cB} & L_{cC} & L_{cA} & L_{cb} & L_{cc} \end{bmatrix} \cdot \begin{bmatrix} i_A \\ i_B \\ i_C \\ i_a \\ i_b \\ i_c \end{bmatrix} \tag{3}$$

In Eq. (3):

$\psi_A, \psi_B, \psi_C, \psi_a, \psi_b, \psi_c$ represent the instantaneous values of the stator and rotor flux linkage, respectively.

Equation (3) simplified:

$$\psi = L \cdot i \tag{4}$$

(c) The torque equation. According to the principle of electromechanical conversion, the torque equation is expressed by the three-phase current and the rotor space position angle:

$$\begin{aligned} T_e =& n_p L_{ms}[(i_A i_a + i_B i_b + i_C i_c)\sin\theta + (i_A i_b + i_B i_c + i_C i_a)\sin(\theta + 120^\circ) + \\ &+ (i_A i_c + i_B i_a + i_C i_b)\sin(\theta - 120^\circ)] \end{aligned} \tag{5}$$

In Eq. (5):

T_e represents electromagnetic torque.
n_p represents the polar logarithm of the motor.
L_{ms} represents maximum mutual inductance flux of interlinking with one phase winding of stator corresponds to the stator mutual inductance.

(d) The motion equation. Ignoring the friction in the motor transmission mechanism and other factors.

$$T_e = T_L + \frac{J}{n_p}\frac{d\omega}{dt} \tag{6}$$

In Eq. (6):

T_L represents load torque.
J represents the moment of inertia of the motor.

3 Coordinate Transformation of the AC Induction Motor

In order to realize the precise control of AC induction motor, the physical model of AC motor is equivalently changed into a model similar to that of a DC motor, and its mathematical model is simplified. So that the problems of the analysis and control of the AC induction motor are simplified.

According to the principle of MT shafting of the AC induction motor, [3] stator currents i_A, i_B, i_C in three-phase stationary coordinate systems A, B and C can obtain the equivalent current i_α, i_β in the two-phase stationary coordinate system α, β by the transformation between three-phase and two-phase stationary coordinate systems of meeting the power invariant requirements.

$$\begin{bmatrix} i_\alpha \\ i_\beta \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{3}{2}} & 0 \\ \frac{1}{\sqrt{2}} & \sqrt{2} \end{bmatrix} \begin{bmatrix} i_A \\ i_B \end{bmatrix} \tag{7}$$

Then the two-phase stationary coordinate system α, β and the two-phase rotating coordinate system M, T are changed, and the equivalent currents i_M, i_T under the MT shafting are obtained.

$$\begin{bmatrix} i_M \\ i_T \end{bmatrix} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} i_\alpha \\ i_\beta \end{bmatrix} \tag{8}$$

In Eq. (8):

φ represents the angle between the α axis of the two-phase stationary coordinate system and the M axis of the two-phase rotating coordinate system.

Therefore, the voltage equation, the flux linkage equation, the torque equation and the motion equation of the AC induction motor under MT shaft system can be obtained.

(a) The voltage equation.

$$\begin{bmatrix} u_M \\ u_T \\ u_m \\ u_t \end{bmatrix} = \begin{bmatrix} R_s + L_s p & -\omega_1 L_s & L_m & -\omega_1 L_m \\ \omega_1 L_s & R_s + L_s p & \omega_1 L_m & L_m p \\ L_m p & -\omega_s L_m & R_s + L_r p & -\omega_s L_r \\ \omega_s L_m & L_m p & \omega_s L_r & R_s + L_r p \end{bmatrix} \cdot \begin{bmatrix} i_M \\ i_T \\ i_m \\ i_t \end{bmatrix} \tag{9}$$

In Eq. (9):

u_M represents the stator M axis voltage, u_T represents the stator T axis voltage.
u_m represents the rotor M axis voltage, u_t represents the rotor T axis voltage, because the rotor winding is short circuited,$u_m = u_t = 0$.
i_M represents the stator M axis current, i_t represents the stator T axis current.
i_m represents the rotor M axis current, i_t represents the rotor T axis current.
R_s represents the resistance of the stator winding, R_r represents the resistance of the rotor winding.
ω_1 represents the stator angular frequency, ω_s represents the rotor angular frequency.
L_s represents the self-inductance between the stator equivalent two-phase windings, L_m represents the self-inductance between the rotor equivalent two-phase windings.

(b) The flux linkage equation.

$$\begin{bmatrix} \psi_M \\ \psi_T \\ \psi_m \\ \psi_t \end{bmatrix} = \begin{bmatrix} L_s & 0_s & L_m & 0 \\ 0 & L_s & 0 & L_m \\ L_m & 0 & L_r & 0 \\ 0 & L_m & 0 & L_r \end{bmatrix} \cdot \begin{bmatrix} i_M \\ i_T \\ i_m \\ i_t \end{bmatrix} \tag{10}$$

In Eq. (10):

ψ_M represents the stator M axis flux linkage, ψ_T represents the stator T axis flux linkage.
ψ_m represents the rotor M axis flux linkage, ψ_t represents the rotor T axis flux linkage.

(c) The torque equation.

$$T_e = n_p L_m (i_T i_m - i_M i_t) \tag{11}$$

(d) The motion equation. The equation of motion remains the same as the Eq. (6).

4 Modeling of Vector Control System Based on Rotor Indirect Field Orientation

Because the indirect magnetic field orientation does not need to observe the actual position of the rotor flux vector, and the orientation is achieved by controlling the slip frequency, it is also called the slip frequency method [4].

The principle of vector control of AC induction motor based on slip frequency control in Fig. 2. The main circuit of the system uses the SPWM three-phase voltage source inverter. Rotational speed is controlled by slip frequency, which according to the stator angular frequency ω_1, the rotor angle frequency ω,the slip angle frequency ω_s of the induction motor satisfy the relation of $\omega_s = \omega_1 - \omega$.

Slip frequency vector control without complex flux detection and coordinate transformation, as long as the premise of ensuring the rotor flux size constant, according to detecting stator current and rotor angular frequency and operating based on the mathematical model of AC induction motor, which can realizes the

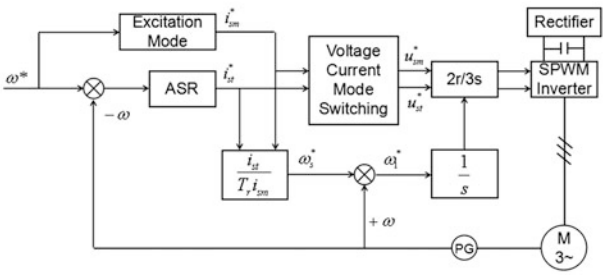


Fig. 2 The principle block diagram of vector control of AC induction motor based on slip frequency control

indirect control based on rotor flux orientation [5]. The following is the basic equation for control:

According to the of the flux linkage equation, the torque equation and The motion equation of the coordinate transformation:

$$T_e = \frac{n_p L_m}{L_r} i_{st} \psi_r \tag{12}$$

So:

$$i_{st} = \frac{n_p L_r}{L_m \psi_r} T_e \tag{13}$$

$$i_{sm} = \frac{T_r p + 1}{L_m} \psi_r \tag{14}$$

$$\omega_s = \frac{L_m}{T_r \psi_r} i_{st} \tag{15}$$

In Eq. (15):

i_{st} represents the torque component of stator current.
ψ_r represents rotor flux linkage.
T_r represents rotor time constant, $T_r = \frac{L_r}{R_r}$.

Because of the excitation and torque components of the stator current are obtained from the vector control equations, but the control system in this paper is a three-phase voltage type inverter, so the corresponding current control is converted to voltage control [6], the transform relationship:

$$u_{sm} = R_s i_{sm} - \omega_1 \sigma L_s i_{st} \tag{16}$$

$$u_{st} = \omega_1 L_s i_{sm} + (R_{s+} \sigma L_s i_{st}) i_{st} \tag{17}$$

In Eq. (15):

i_{sm} represents the excitation component of the stator current.
σ represents magnetic leakage factor.
u_{sm} represents the excitation component of the stator voltage.
u_{st} represents the torque component of stator voltage.

The voltage control signal of the SPWM three-phase voltage inverter is obtained by coordinate transformation, which controls the output voltage of the inverter, so that the AC induction motor drives the load to run.

5 Simulation of AC Induction Motor Control System Based on MATLAB/Simulink

The slip frequency control of the AC induction motor speed control system is mainly composed of the main circuit model and the control circuit model. Among them, the main circuit model includes IGBT inverter circuit, motor module and measurement module. The control circuit model is mainly composed of speed PI regulator module, ω_1 generation module, u_{sm}, u_{st} generation module, coordinate conversion module. The functions and algorithms of each part are as follows.

5.1 Main Circuit Model

The main circuit model of the control system is composed of three parts: IGBT inverter circuit, motor module and measurement module. The IGBT inverter circuit with self-shutting IGBT as the converter device; the motor module selects the Asynchronous Machine SI Units motor from Simulink; the measurement module chooses rotor current, rotor flux, stator current, stator flux, rotor angular frequency and electromagnetic torque as measurement items.

5.2 Control Circuit Model

(a) Speed PI regulator module. In Fig. 3, the input of the PI regulator is the difference between the motor rotor angular frequency ω^* and the actual rotor angular frequency ω of the motor. And the output of the PI regulator is limited to obtain the torque component of the stator current i_{st}^* [7].

(b) ω_1 generate modules. In Fig. 4, According to $\omega_s + \omega = \omega_1$, the stator angular frequency ω_1 can be obtained.

(c) u_{sm}, u_{st} generation module. In Fig. 5, u_{sm}, u_{st} can be calculated according to the Eqs. (16) and (17).

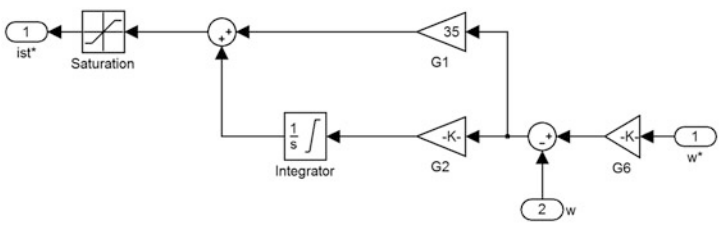


Fig. 3 Speed PI regulator module

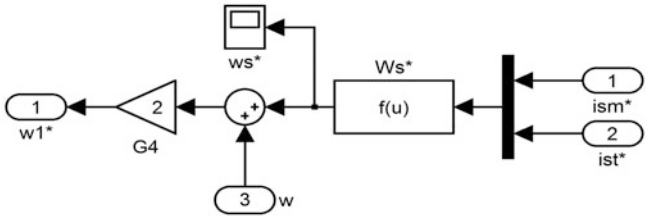


Fig. 4 ω_1 generate modules

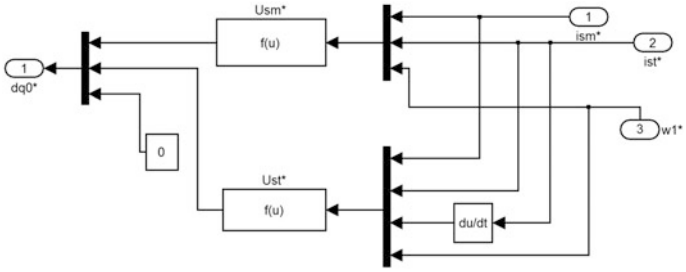


Fig. 5 u_{sm}, u_{st} generation module

(d) Coordinate conversion module. In Fig. 6, phase θ can be obtained by the Integral adjustment of the Stator angular frequency ω_1^*, and it and voltage amplitude under the two-phase rotating coordinate system (dq0) are used as inputs to the module. Finally, the module outputs three-phase stator voltage u_a, u_b, u_c under the three-phase stationary coordinate system (abc).

5.3 *Simulation Model of Vector Control Speed Regulation System Based on Rotor Indirect Field Oriented*

Simulation model of vector control speed regulation system based on rotor indirect field oriented in Fig. 7.

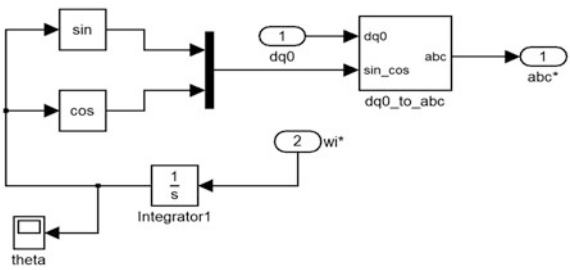


Fig. 6 Coordinate conversion module

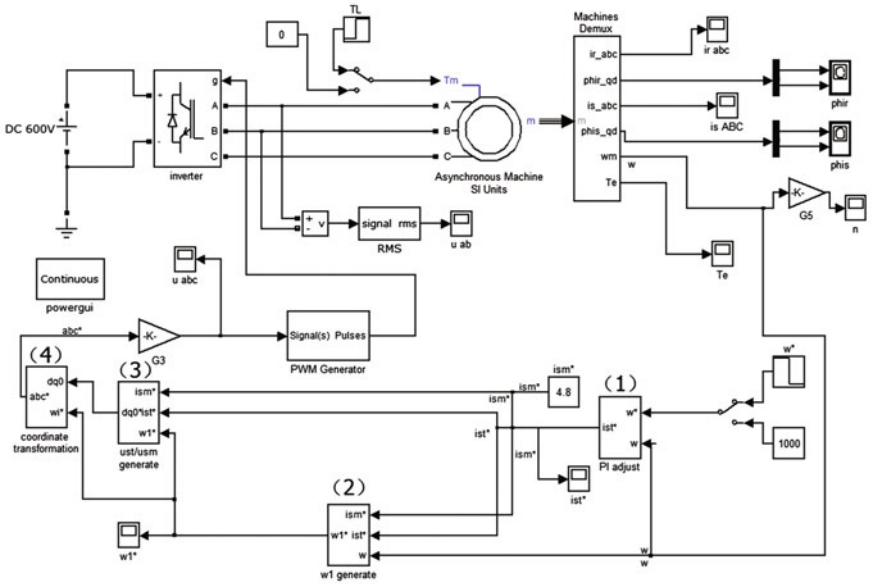


Fig. 7 Simulation model of vector control speed regulation system based on rotor indirect field oriented

5.4 Parameter Setting of the Induction AC Motor

The simulation motor selects a brand low speed pure **electric** vehicle driven induction AC motor, and its main parameters in Table 1.

Table 1 The simulation motor main parameters

Sign	Parameters	Value
p_n	Rated power	3730 W
U_n	Rated voltage	460 V
f_n	Rated frequency	60 Hz
R_s	Stator windings	1.115 Ω
L_{1s}	Stator leakage inductance	0.006 H
R_r	Rotor windings	1.083 Ω
L'_{1r}	Rotor leakage inductance	0.006 H
L_m	Stator and rotor mutual inductance	0.007 H
J	Moment of inertia	0.189 kg·m^2
U	Inverter DC power supply	600 V
F	Coefficient of friction	0
n_p	Pole-pair number of the motor	2

5.5 Simulation Results and Analysis

In order to simulate the real situation of the vehicle during the driving process, we given the initial speed of the motor n^* = 1000 r/min when the motor is empty. When the 0.45 s is started, the load T_L = 50 Nm is suddenly loaded with the motor; and when the 0.7 s is started, the speed n^* mutation to −1000 r/min.

The motor speed, the stator three-phase current, the inverter output three phase line voltage (motor input three-phase line voltage) and the motor electromagnetic torque response curve of the simulation process in Fig. 8.

In the process of starting, the speed increases rapidly, the inverter output three-phase voltage also gradually increased, the electromagnetic torque also fluctuates rapidly after rapid rise, and the stator current fluctuates, range of fluctuation is smaller than that of speed, voltage, and torque.

At 0.17 s, after the speed is slightly overshoot (the maximum overshoot rate is 0.20%, as shown in Fig. 9), the speed stabilized at the no-load speed 1000 r/min. At this point, the stator current is also decreased to the no-load current, the output voltage of the inverter is correspondingly decreased, and the electromagnetic torque remains relatively stable; At 0.45 s, the motor load, current, voltage, torque are rapidly rising, the torque immediately rose to a given torque of 50 Nm, and after the speed decreases slightly (the maximum change rate of 0.25%, as shown in Fig. 10), the speed stabilized at the no-load speed; At 0.7 s; the motor begins to reverse, speed drops immediately, until 0.97 s, after the speed is slightly overshoot

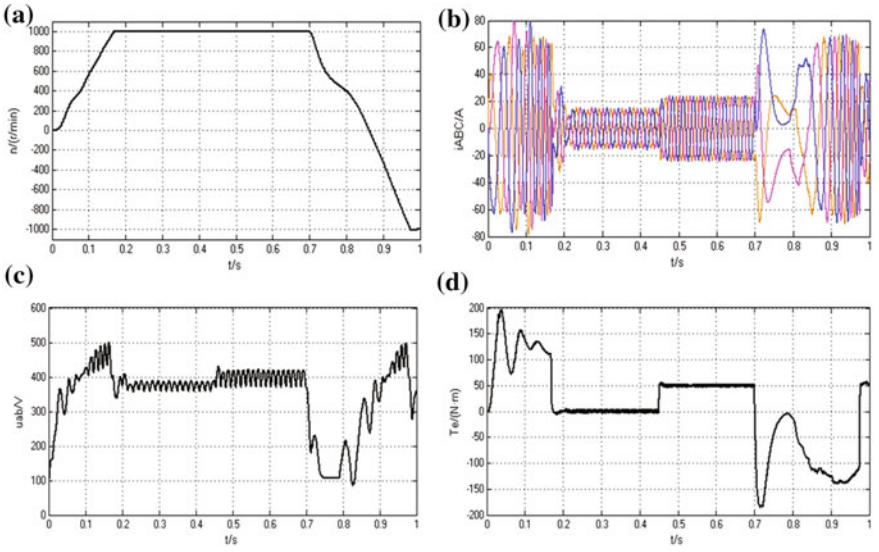


Fig. 8 Response curve **a** Speed response **b** Stator three-phase current response **c** Motor input three-phase voltage response **d** Electromagnetic torque response

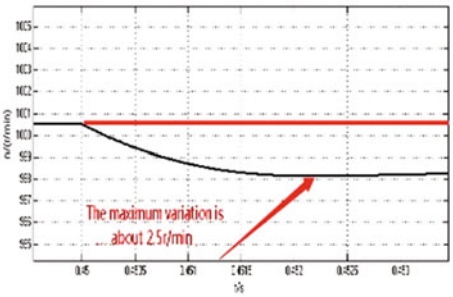


Fig. 9 (0.17 s) overshoot of speed

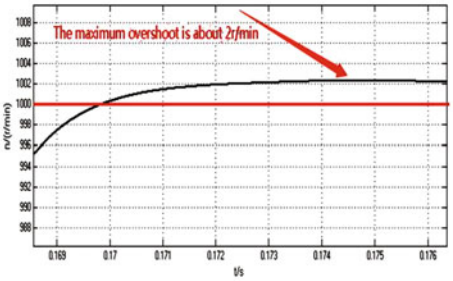


Fig. 10 (0.45 s) slightly changed of speed

(maximum overshoot rate is 0.50%, as shown in Fig. 11), the speed stabilized at reverse speed −1000 r/min until stop. The current also changes rapidly, it changes after a disorder of about 0.15 s, and it also maintains good three-phase sine change until stop. The voltage and torque decrease rapidly after the motor starts reversing, and after a period of great change, they maintain relatively small amplitude changes before the motor is about to stop.

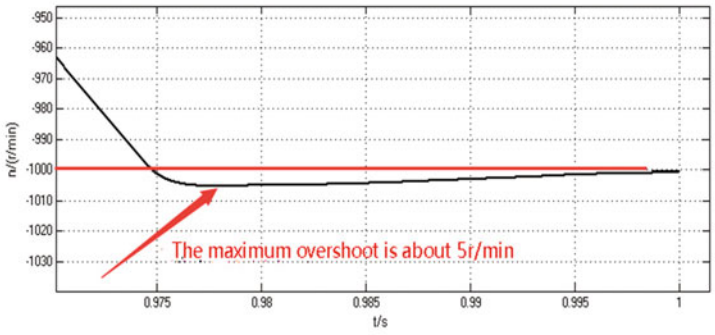


Fig. 11 (0.9 s) overshoot of speed

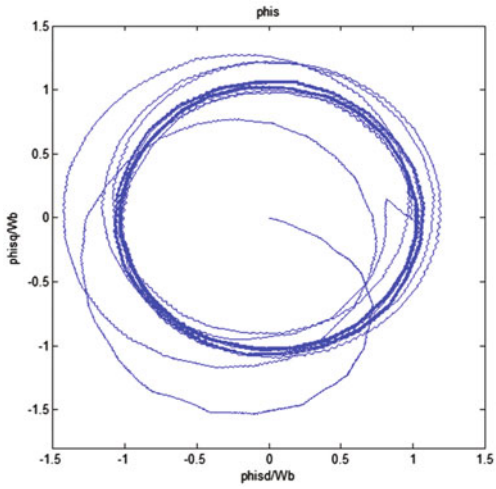


Fig. 12 Stator flux linkage trajectory of a motor without inversion

In Fig. 8d, the fluctuation of the motor torque is due to the difference between the output voltage of the three-phase voltage inverter and the given DC supply voltage. In order to reduce the fluctuation, the algorithm of the inverter can be improved theoretically to reduce the fluctuation [8].

If the motor is not turned back at 0.7 s, the motor still keeps the forward speed of 1000 r/min. After simulation, the stator flux linkage of the motor is obtained, as shown in Fig. 12. The analysis can be made that when the motor starts in zero state, the magnetic field establishment needs a process, and the magnetic field is not all regular change in the process. In the first 0.18 s after the start of the motor, the stator flux linkage trajectory is irregular shape, and then the trajectory is regular circle. In combination with Fig. 8d, the motor torque changes substantially within 0.18 s of the start of the motor, and also due to an irregular change in the magnetic field.

References

1. Yang NM (2016) Control strategy for motor drive system of electric vehicle. Ieej Trans Electr Electron Eng 11(3)
2. Pellegrino G, Bojoi RI, Guglielmi P (2011) Unified direct-flux vector control for AC motor drives. IEEE Trans Ind Appl 47(5):2093–2102
3. Faugeras OD, Bras-Mehlman EL, Boissonnat JD (1990) Representing stereo data with the Delaunay triangulation. Artif Intell 44(1–2):41–87
4. Xia XK, Zhang JZ, Cheng M (2009) Research on slip frequency vector control scheme of asynchronous motor. Micromotors 6
5. Koubaa Y, Boussak M (2013) Rotor resistance tuning for indirect stator flux orientated induction motor drive based on MRAS scheme. Eur Trans Electr Power 15(6):557–570

6. Zhang K, Liu Z, Zhang L et al (2012) Modeling of high-speed motorized spindle and design and simulation research of vector controller based on slip frequency control. J Shenyang Jianzhu Univ 28(6):1114–1120
7. Zhu J, Chen L, Xiong T, et al (2012) Modeling and research on slip frequency vector control system of electric forklift. Mod Manuf Eng 4
8. Wang X, Yang Y, Liu W (2011) Simulation of vector controlled adjustable speed System of induction motor based on Simulink. International Conference on Computer Science and Service System. IEEE

Vision Based Quality Control and Maintenance in High Volume Production by Use of Zero Defect Strategies

Håkon Raabe, Odd Myklebust and Ragnhild Eleftheriadis

Abstract Quality inspection in high volume manufacturing often depends on manual decision-making processes and predictive condition based maintenance. Predictive maintenance as a part of defect detection related to indicators and environmental impact in production development. Use of Zero Defect (ZD) methods or Cyber Physical Systems (CPS) will give great value to reduction of scrap rates, fault and defects. Scrap rates is often a huge problem due to different conditions in the production system, and will often cause impact on the delivery to costumer. To provide a self-correcting system by use of machine control data and augmented reality (AR) solutions, input to continuous maintenance tasks on the shop floor can be assisted. CPS, vision based inspection or AR solutions in the production environment provides the opportunity to detect fault in order to prescribe a possible maintenance solutions. The vision inspection as decision-making system may give feedback on detected problems, like contrast of parts, change in geometry, material or smutch on product or on the palette in production. This makes it possible to reduce interoperability and to make self-adjustments in production lines by use of CPS as a system for collaborating computational entities connected to services.

Keywords Zero defect manufacturing · Maintenance · Cyber physical systems

H. Raabe (✉)
Sintef Raufoss Manufacturing, NMK, Borgundvegen 340, Ålesund 6009, Norway
e-mail: Hakon.Raabe@sintef.no

O. Myklebust · R. Eleftheriadis
Sintef Raufoss Manufacturing, S.P. Andersensv. 5, Trondheim 7031, Norway
e-mail: Odd.Myklebust@sintef.no

R. Eleftheriadis
e-mail: ragnhild.eleftheriadis@sintef.no

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_43

1 Zero Defect Strategies in Multi Stage Production

Rapid development in high volume production, with combination of different materials and complex components creates a need for zero defect strategies and solutions for maintenance operations to ensure quality and reliability on the product. In a high volume production the focus on customer demand is not the most important; however, an inflexible production often give a higher amount of fault and product which can give a higher cost on the finished product. With self-adapting and visual solutions, it is possible to maintain the manufacturing process, to save cost of failures on products and assets as well as correcting faults before they occur to achieve better and more reliable products. With use of Zero Defect Manufacturing strategies (ZDM) [3, 4], Industry 4.0 and Cyber Physical System [7, 11] it is possible to boost production line effectiveness, using methods of effective decision-making. Both Deming and Kaplan & Norton have captured the important ways of how to find the right Key Performance Indicators (KPI) for excellent decision making in an organization. For a high volume company this is even more important since faults will propagate and increase in a more escalating way [5, 17].

We will briefly mention the use of Lean, TPM (Total Productive Maintenance) and OEE (overall equipment efficiency) [14, 16] as the most frequently used maintenance strategies in high volume companies. In this paper, we will present one industrial case, which is in progress on an industrial site. Early investigations provide very good feedback on how to use Augmented Reality (AR) to predict when an asset needs to be changed. This method is based on an image projection, which make it possible to tell when the part is in right position or when to change due to dirt, dust or oil spill on the palettes in the environment. By implementing a visual tracking of datasets, fault can be detected in the pictured grayscale on the palettes, so that they will be changed and maintained in due time. The decision making process rely on the use of camera(s), algorithms and technologies for data processing.

2 Quality Approach in Modern Manufacturing

2.1 Different Quality Methods

In high cost and complex production and high volume product, it is necessary to hold a large focus on maintenance operations due to increase in cost, on the material but also maintenance cost on the services and asset. Production quality is a discipline that combines quality, production equipment and maintenance tools. This to maintain the throughput and the service level of conforming parts under control and to improve and maintain them over time, with minimal waste of resources and materials [2].

Quality control methods are coming from different frameworks during the last century approaches like TQM, Lean, TPM, 5S and Six Sigma. These are mixed approaches or methods where interconnected views are used. Some of the methods are qualitative, others are quantitative approaches [15]. To continuously improve or specify a quality problem in readability processes, you often need to use a mixed approach. A beyond six sigma approach is used for finding quality improvement parameters, and these goals fulfil process and product expectation of the costumer [5]. There are 3 dimensions of how to understand six sigma principles:

- To know your costumers needs
- To split and simplify processes
- The reduction of variation.

Through including sensors in processes, machines, tools and embedded systems, it is possible to handle Big Data with AI systems and new statistical and non-statistical approaches in the different quality cases. SPC for small batch production, Zero Defect Manufacturing and plant learning solutions (real-time and off-line corrections) [6].

2.2 Zero Defect Manufacturing

The ZDM philosophy is a different way of thinking and it is focusing on value creation, compared to focus on cost. We know that most companies have a lot of aggregated data and many do not know what to do and how to use all their available information. The literature on machinery diagnostics and prognostics is diverse primarily due to a wide variety of systems, components and parts and are often referred to as a service-oriented technology or Cyber Physical Production Systems. In a multistage process system, cost effective quality control and ZDM is introduced by manufacturing system strategies that by mapping the occurrence of the defects by use of prediction and prognosis system and an integrated models, effective can adopt Big Data by use of sensors and then optimize quality control loops that can be systematically visualized analyzed and improved [6].

The ZDM quality mapping can by collecting the right data use preventative maintenance as performed in real time while the production equipment is still working. This new era of maintenance in production without unexpectedly breakdowns and stops, is a new technology area for manufacturing solutions [13].

Adaption of ZDM philosophy was necessary with focus on some few, but important quality feedback loops. The collected data can be tied to digital models created on the corresponding production systems. Components and sensors provide a self-descriptive knowledge base for various data mining tasks that optimizes existing systems [6].

3 Cyber Physical System and Maintenance

3.1 The Maintenance View

Usually most high volume manufacturers use TPM philosophy from Nakajima [14], which looks at TPM as a tool to maximize OEE. It is an effective way of analyzing the efficiency of a single machine or an integrated manufacturing system, since it measures the losses by function of availability, performance rate or quality rate in production.

TPM is using a proactive and cost effective approach to maximize equipment effectiveness, using principles of teamwork, empower, "zero breakdowns" and "zero defects". However, the use of OEE have some weak links and can be criticized due to lack of constraint in the production and aggregate of metrics [10]. This negative input of the OEE can provide an increase in the numbers of errors and lead to higher scrap rates and defective products due to degradation affected by environmental conditions in the assembly line on such as vibration, temperature, oil deposit and dust.

This mean that a high efficiency not always is the best solution for a high volume manufacturer, if the design, material or other incidents influence the assembly processes. In continuous maintenance, degradation and assessment techniques is usually; visual inspection, dye penetrant inspection, magnetic particle inspection, ultrasonic testing and thermography just to mention some of them. This is often a part of the product assembly or in the end of a process. Checking degradation of a component in service using non-destructive evaluation techniques and automating the assessment processes is the trend, like automated repair and robot guided rework, rapid manufacturing of spare parts or additive manufacturing [15].

3.2 Cyber Physical Production Systems and Life Cycle Processes

The embedded systems of CPS are able to bridge the gap between the real and the virtual world by connecting smart objects with factory information systems, making all its actors interacting with each other across the entire value chain [7, 8]. The combination of embedded low power and high performance computing (HPC) enables autonomous generation of a huge amount of data, which is becoming a crucial aspect in every industry's sector to meet the increasing complexity in costumers demand, both the product and factory life cycle. In the first case by providing real-time information about last stages of the life cycle (e.g. de-manufacturing, aftersales, etc.) [13].

The term of CPS refers to a new generation of systems with integrated computational and physical capabilities that can interact with humans through many new modalities in the production system. The ability to interact with, and expand

the capabilities of the physical world through computation, communication, and control is a key enabler for future technology developments. In combination with a model-based and linked-data approach, this can help to reduce the number of faults and the complexity of interfaces for data access. The new technological solutions and techniques for collecting data and information from the MES system (manufacturing execution system), establish a new area in the quality system for integrated real time monitoring, and predictable control of vital process parameters that can be monitored in real-time [9].

The scope of CPS is broad and bring also the human factor and knowledge capturing to be taken care of, and both the product and the factory life-cycle which will be impacted by ZDM. Just for considering the sustainability aspect during the initial phases of the product/service concept and the design processes Fig. 1 [13] show how maintenance and enterprise models is used to structure the whole ZDM concept in the same way as CPS and CPPS is done today [12]. When the production of today manufacture in accordance with the production line, we know that tomorrow's manufacturing systems will be much more decoupled, fully flexible and highly-integrated. Preventive maintenance not only enables high service, it will also affect the quality of parts produced.

Augmented reality (AR) is what you see right now and give a kind of training, supplemented by sensor techniques like sound, video or graphics or GPS data, in real-time. Augmented reality brings out the components of the digital world into a person's perceived real world. Hardware components for augmented reality are processor, display, sensors and input devices [1]. Modern mobile computing devices like smartphones and tablet computers contain these elements, which often include a camera and MEMS sensors such as accelerometer, GPS, and solid-state compass, making them suitable for AR platforms. Optical handheld devices or projector systems can be used as devices for recognition of surface or display the different colors of a palette or a surface (Fig. 2).

4 Machine Vision Inspection and Maintenance in a Real Case

The aim for this research case was to ensure a zero defect strategy on a composite product in the automotive supply market. The company operates in industries where there are high demands on the quality of products and where zero-defect production

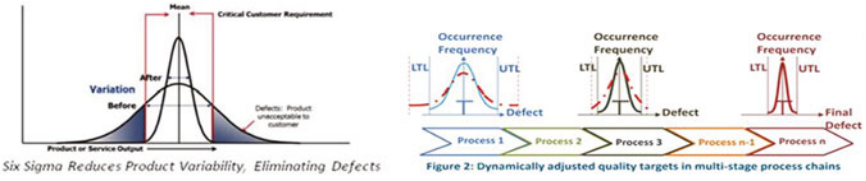


Fig. 1 ZDM best practice and benchmark [5]

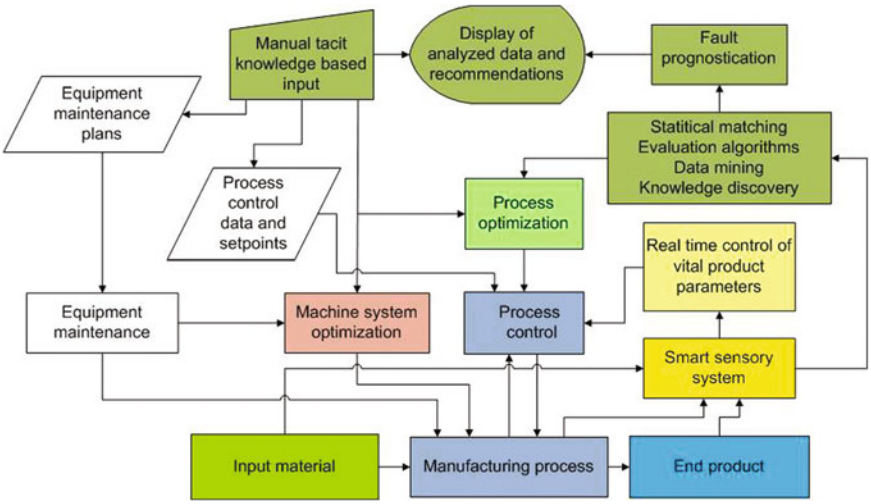


Fig. 2 Zero defect manufacturing framework [6]

is of great significance for the customer. There are no acceptance for wrong deliveries of the product with regard to the expected quality, and faulty products can cause serious consequences that at the worst can lead to loss of life. The product have a failure rate within the tolerances of normal mass production, however, the achievement was to reduce the scrap rate close to zero by improved control of material handling and processing, and integrated and controlled use of resources for maintaining of the production and the assembly line. The product envisions contributing to reduction of internal scraps of 1 percentage point; however, each link is composed of 5–10 internal components. We can see that added up, even small effects would reduce the time/cost per unit and make big savings over a year.

During the research period, we have discovered different challenges and incidents in the working environment that lead us to test a strategy of using visual AR techniques to predict fault on the manufacturing operations in process. The company have separate data collection systems that collect information from the individual processes, but a combination and composition of data from all elements of the process is not yet resolved. The thought behind the system is to use ZDM and CPS in a combination with an algorithm that collect all the visual images on each critical points observed from a maintenance view. Multistage manufacturing processes is quite complex to understand and how a defect generates in a certain process stage and propagates to the next step is quite a difficult task. Advanced software tools to analyze stage correlation is still not available to buy, but can be developed for each production site.

The system will then be able to analyse and sort out defects in each necessary step in the production process. This to handle errors and discrepancies and to ensure that planned changes and justifications will be correctly implemented in the production. In this way, the production process can be optimized and scrap of products

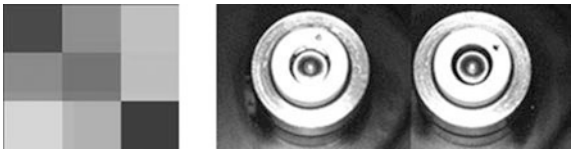


Fig. 3 The difficulties of distinguishing between different grey scales

can be avoided. A statistical multivariate analysis is allowing us to move the decision point further upstream and we can create a multidimensional picture of the reality in real-time. Multivariate systems has often a large amount of data, and analysis can often be influenced by the results. Broad product knowledge and analytical skilled operators is needed to get the right result. Camera systems of today is currently insufficient regarding image interpretation. One way to analyze images of today is using artificial intelligence. The algorithms work in parallel and make decisions based on picked parameters from the production line. However, such algorithms do not follow the analysis backwards, so you need a statistical basis to predict this before it become an optimal tool. By use of this imaging tool and logged data we can perform preventive maintenance while the equipment is still working, and it will be possible to make maintenance operations and avoid unexpected break-down of the processes (Fig. 3).

A decision support system that can handle this analysis when the parts' position are visualized via scanning, based on a grayscale that can give feedback on faults like dust on a palette, the thickness of the container, or correct positioning of an O-ring, will give a great benefit to value creation in the production. Manufacturing in a multi stage assembly line with use of machine vision technology to detect faults in production and assembly lines, looks like a promising methodology with the findings from this study.

5 Conclusion

This paper has reviewed the investigation done in a company producing safety critical plastic composite parts, which needed a solution for real-time quality control of high volume commodity production due to large cost of maintenance operations. The case has measured imaging data and different algorithms to optimize production processes that need to produce according to customer requirements in the market. The challenge has not been to collect enough data, but to get all the right decisions about what should be measured, at the appropriate time, and how, in order to have an adequate and accurate decision-making for avoiding "information overload." The case has shown how high process variability, and focus on the value to the customer that can provide a first-time-right-quality-strategy or a Zero Defect Manufacturing strategy [2, 5]. A strategy that gives high productivity and reduces scrap and waste that will benefit the company as well as the customer. The study has also showed that production quality can be improved by control systems and improvement of parts focusing minimization of waste of resources and material consumption.

References

1. Azuma RT (1997) A survey of augmented reality. Presence 6(4):355–385
2. Colledani M et al (2014) Design and management of manufacturing systems for production quality., journal homepage: http://ees.elsevier.com/cirp/default.asp, CIRP Ann—Manufact Technol 63:773–796
3. Crosby P (1969) Quality is free. McGraw-Hill, New York
4. Deming E (1982) "Out of the crisis", quality, productivity and competitive position. MIT Press, Cambridge
5. Eleftheriadis R, Capellan A, Myklebust O (2016) Continuous improvement and benchmarking to achieve zero defect manufacturing (ZDM). Atlantis Press, Shanghai, IWAMA 2014
6. IFaCOM (2015) IFaCOM final summary report, available via http://cordis.europa.eu/project/rcn/101390_en.html. Trondheim, Norway, European Commission, R & D Department: European Commission, http://cordis.europa.eu/project/rcn/101390_en.html
7. Juran JM (1979) Quality control handbook. McGraw Hill, New York, pp 5–12
8. Kaplan RS, Norton DP (1996) Translating strategy into action. The balanced score card. Harvard Business School Press, Boston
9. Lee EA (2008) Cyber physical system; design challenges. San Francisco: University of California at Berkeley, technical report no. UCB/EECS-2008–8
10. Miyake DI, Enkawa T, Fleury AAC (1995) Improving manufacturing systems performance by complementary application on just-in-time, total quality control and total productive maintenance paradigms. Total Qual Manage 6(4):345–363
11. Monostori L (2014) Cyber-physical production systems: roots, expectations and R&D challenges. Canada: s.n., 2014. In: Proceedings of the 47th CIRP conference on manufacturing, Elsevier, Amsterdam
12. Monostori L et al (2016) Cyber-physical systems in manufacturing. CIRP Ann—Manufact Technol 65(2016):621–641
13. Myklebust O (2013) Zero defect manufacturing; a product and plant oriented lifecycle approach. In: CIRP Procedia 12, 8th CIRP conference on intelligent computation in manufacturing engineering, Napoli, pp 246–251
14. Nakajima S (1988) Introduction to total predictive maintenance (TPM). Productivity Press, Cambridge
15. Roy R et al (2016) Continuous maintenance and the future—foundations and technological challenges. CIRP Ann—Manufact Technol, 667–688
16. Wang K-S (2013) Towards zero defect manufacturing—a data mining approach. Adv Manufact, 62–74
17. Womack JP, Jones DT, Roos D (1990) The machine that changed the world. Rawson Associated Scribner, New York

A RFID Based Solution for Managing the Order-Picking Operation in Warehouse

Haishu Ma, Jinghui Yang and Kesheng Wang

Abstract As an essential component in the supply chain, warehouse links the production in the upstream and distribution in the downstream. The efficiency of the whole supply chain is significantly affected by the performance of warehouse operations. The order picking process has been considered as the most time consuming and costly warehouse operations in warehouse management. This paper proposes a RFID based solution for managing the order picking operation in the warehouse. The solution first calculates the exact locations of forklifts tagged with RFID by means of k-nearest neighbor method. Then an integer-linear mathematical programming model is constructed to formulate the shortest travel distance of forklifts. The proposed solution is tested in simulations. The result demonstrates that our method is able to generate the optimum pick-up sequence for the appropriate forklift. In doing this the objective of improving the effectiveness and efficiency of order picking operations is achieved.

Keywords RFID · Warehouse operation · Order-picking · Route optimization

1 Introduction

Due to the highly competitive global business environment, the functions of warehouse in a supply chain are no longer confined to keeping a large amount of stock. Instead, customer orders with high product varieties in small quantities are

H. Ma (✉) · K. Wang
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology (NTNU), NO 7491 Trondheim, Norway
e-mail: Haishu.ma@ntnu.no

K. Wang
e-mail: Kesheng.wang@ntnu.no

J. Yang
College of Engineering, Shanghai Polytechnic University, Jinhai Road 2360, Pudong, Shanghai, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_44

often received by the logistics service providers. The fulfillment of customer orders in the warehouse is challenging since it is essential to satisfy increasing customer demand regarding responsiveness, cost effectiveness and flexibility.

The main operations that almost every warehouse needs to manage involve receiving inbound SKUs from suppliers, storing the SKUs, receiving orders from customers, retrieving the requested items and outbound shipping [1]. The performance of these operations affects the productivity of the warehouse, even the whole supply chain. The order picking process has been considered as the most time consuming and costly warehouse operations in warehouse management [2]. Efficient product location management serves as the key to reducing product searching time and traveling time during the order picking operation. The recent research in the warehouse shows that more than 70% of the time was spent on searching the designated forklifts and travelling to the shipping dock for loading [3]. To improve productivity in warehouse, it is crucial to obtain timely location information of all resources for accurate decisions in warehousing processes. The RFID-based system in product location management has been widely studied [4, 5]. Poon et al. [6] established a RFID case based logistics resource management system to improve visibility and productivity of the order picking process. But the locations of warehouse resources were determined by triangular localization scheme which was not accurate. Chow et al. [7] proposed the RFID-RMS system to maximize the utilization of warehouse resource and enhance the work efficiency. However, the forklift was located by ultra-wide band technology (UWB), which has high implementation cost. Choy et al. [8] presented a RFID based storage assignment system to provide decision support for enhancing the efficiency of order picking. Unlike many existing situations, RFID tags were attached to products at the item level instead of at the pallet level, which is not practical and unnecessary.

This paper proposes a RFID based solution to increase the efficiency of order-picking operation in warehouse. With the help of RFID, the feature of real-time and automatic data collection is achieved, which also facilitates the construction of an effective localization scheme to pinpoint the exact locations of forklifts. Moreover, a route optimization model is developed using the collected RFID data to handle the order-picking operations. The rest of this paper is organized as follows: Sect. 2 illustrates the solution for the order-picking problem. The performance of the proposed solution is evaluated in Sect. 3. Finally, the conclusions are given in Sect. 4.

2 Methodology

The aim of this solution is to enhance order-picking operations by means of tracking, positioning, and optimizing resource utilization. Figure 1 illustrates the architecture of the RFID based solution, which mainly comprises three components, namely data collection, localization scheme, and route optimization. The passive RFID tag is attached on the items like pallets and forklifts to record their identities

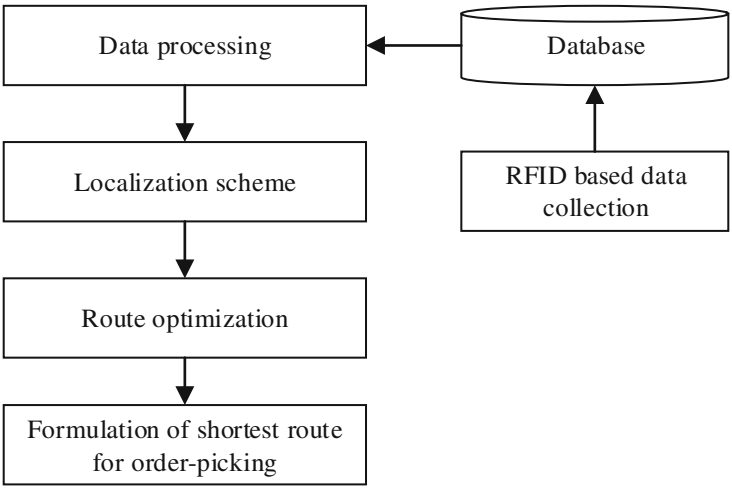


Fig. 1 Architecture of the RFID based solution for the order-picking operation

and exchange data with RFID reader. The reader is attached to the fixed facility in the warehouse, such as the main entrance, storage racks and the dock door, to send and receive the backscattered signals from RFID tag. The reader antennas are capable of reading hundreds of tags within their reading range. The collected data is stored in the centralized database. The received signal strength collected by the reader antenna is used in the localization scheme.

2.1 Localization Scheme

K Nearest Neighbor (KNN) is a well-known RSSI based method and has been widely used in location estimation [7, 10]. KNN employs the RSSI difference between a pair of tags to identify the nearest reference tags and then estimate the target's position. The algorithm is shown as follows:

n	number of reader antennas
m	number of reference tags
$S = (s_1, s_2, \ldots, s_n)$	signal strength vector of target tag
$\theta = (\theta_1, \theta_2, \ldots \theta_n)$	signal strength vector of reference tag

Step 1: Determine the k nearest reference tags.

First we calculate the Euclidean distance in signal strengths which is computed as $E_j = \sqrt{\sum_{i=1}^{n} (\theta_i - s_i)^2}$. For m reference tags, the Euclidean distance vector is $E = (E_1, E_2, \ldots, E_m)$. Then we sort the vector in ascending order. Since the nearer the reference tag is to the target tag, the smaller the E_j is supposed to be. The

k nearest reference tags are determined as the first three tags in the Euclidean distance vector in signal strengths.

Step 2: Determine the weights of the k nearest reference tags

The weight w_i depends on the E_i value of each reference tag. It is calculated as,

$$w_i = \frac{1/E_i^2}{\sum_{i=1}^{k} 1/E_i^2} \tag{1}$$

Step 3: Determine the coordinates of the target tag

The unknown target tag's coordinate is estimated by $(x, y) = \sum_{i=1}^{k} w_i(x_i, y_i)$, where (x_i, y_i) is the coordinate of the reference tag.

2.2 Route Optimization

To calculate the travel distance and the picking sequence for material handling equipment, an integer-linear programming model is constructed with reference to [6, 9]. The parameters and variables used in the route optimization module are shown below.

p	number of forklifts
m	index of forklifts (m = 1, 2 … p)
s, t	index of SKUs (s, t = 1, 2 … q)
c	number of customer orders
a	number of cross aisles, where $a = (a_1, a_2 \ldots a_n)$
S_c	number of SKUs in customer order(c)
d_{st}	distance between two successive SKUs locations. (If S = 0, the location of forklift(m) is the starting point)
Z_{st} = 1	if forklift(m) goes from SKU(s) to SKU(t); 0 otherwise
U_s	pick sequence of SKUs (u_s = 1, 2, … q)

The integer-linear programming model:

$$\text{Minimize} \sum_{s=0}^{q} \sum_{\substack{t=1 \\ t \neq s}}^{q} d_{st} z_{st} \tag{2}$$

Subject to:

$$\sum_{s=0}^{q} z_{st} = 1 \quad \text{for t} = 1, 2, \ldots \text{q}; \quad s \neq t \tag{3}$$

$$\sum_{t=1}^{q} z_{st} \leq 1 \quad \text{for } s = 0, 1, 2\ldots, \text{q}; \quad s \neq t \tag{4}$$

$$u_s - u_t + q z_{st} \leq q - 1 \quad \text{for } s, t = 1, 2, \ldots \text{q}; \quad s \neq t \tag{5}$$

where:

$$d_{st} = |X_s - X_t| + \min \begin{cases} |Y_s - Y_{a1}| + |Y_{a1} - Y_t| & \textit{if cross Aisle } 1 \\ |Y_s - Y_{a2}| + |Y_{a2} - Y_t| & \textit{if cross Aisle } 2 \\ \vdots & \vdots \\ |Y_s - Y_{an}| + |Y_{an} - Y_t| & \textit{if cross Aisle } n \end{cases} \tag{6}$$

Coordinates of forklifts: $\left[X_{(1,2,\ldots p)}, Y_{(1,2,\ldots p)}\right]$; Coordinates of SKUs: $\left[X_{(1,2,\ldots q)}, Y_{(1,2,\ldots q)}\right]$; Coordinates of aisles: $\left[X_{(a_1,a_2,\ldots a_n)}, Y_{(a_1,a_2,\ldots a_n)}\right]$.

The goal of objective Eq. (2) in the programming model is to optimize the travel distance and picking sequence for the order picking operation.

Constraint (3) ensures that only one type of SKU is picked before the forklift goes to the next SKU storage location (t).

Constraint (4) ensures that at most one type of SKU is picked after the forklift leaves the SKU storage location(s), in case that SKU storage location(s) is the last type of SKU in customer order(c).

Constraint (5) is applied to avoid the occurrence of sub trip. As Z_{st} = 1 means that forklift goes from SKU storage location(s) to storage location(t), the sequence is u_s then u_t.

3 Simulation

To facilitate the understanding the route optimization model, an example of optimizing the picking sequence for order picking is illustrated below. Assume the layout of the warehouse is shown in Fig. 2 and the coordinates of forklifts are calculated using the localization scheme. There are four SKUs, two forklifts, and four Cross aisles. The shortest route can be obtained after these parameters are substituted into the mathematical programming model.

To solve the optimization problem above, the standard Mixed-integer linear programming solver in MATLAB is applied [10]. The minimum of the problem is specified by,

$$\text{Min} f^T x \text{ subject to} \begin{cases} x \textit{ are integers} \\ A \cdot x \leq b \\ Aeq \cdot x = beq \\ lb \leq x \leq ub \end{cases} \tag{7}$$

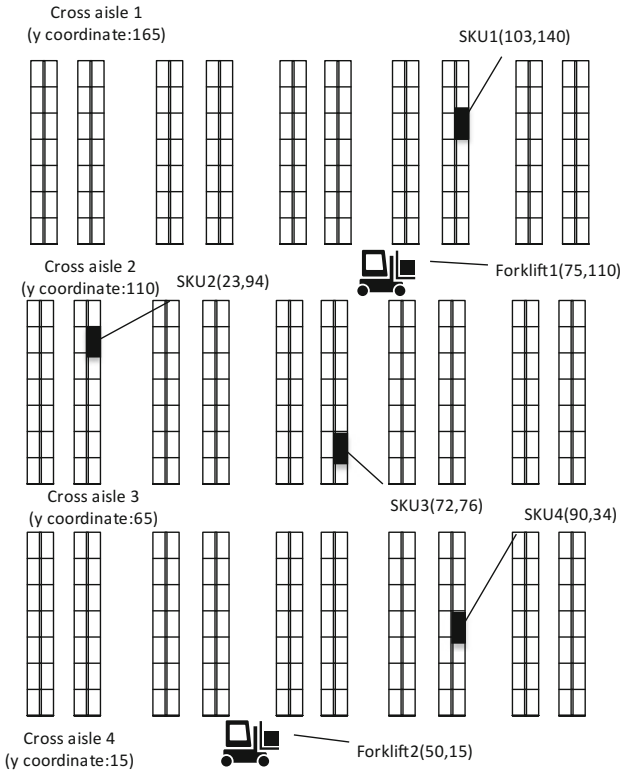


Fig. 2 Layout of warehouse

where f is the coefficient vector. A and b are the linear inequality constraint matrix. Aeq and beq are the linear equality constraint matrix. The lower and upper bounds of x are lb and ub respectively.

The shortest travel distance of forklift m_1 is 344 m and the order picking sequence is $SKU4 \rightarrow SKU3 \rightarrow SKU2 \rightarrow SKU1$.

The shortest travel distance of forklift m_2 is 326 m and the order picking sequence is $SKU1 \rightarrow SKU4 \rightarrow SKU3 \rightarrow SKU2$.

As the travel distance of forklift m_2 is shorter than forklift m_1, forklift m_2 is selected to perform the order picking following the defined SKUS picking sequence.

4 Conclusions

This paper presents a RFID based solution for managing the order-picking operation in the warehouse. The incorporation of KNN localization scheme and route optimization model helps the warehouse to improve the working efficiency. An example of optimizing the picking sequence for order picking is illustrated to facilitate the understanding the route optimization model. Based on the proposed mathematical model, the shortest travel distance and optimum pickup sequence for the appropriate material handling equipment are determined. Nevertheless, the proposed is only verified in simulations. Future work will conduct field test in a warehouse to evaluate the performance of the model.

References

1. Gu J, Goetschalckx M, McGinnis LF (2007) Research on warehouse operation: a comprehensive review. Eur J Oper Res 177(1):1–21
2. De Koster R, Le-Duc T, Roodbergen KJ (2007) Design and control of warehouse order picking: a literature review. Eur J Oper Res 182(2):481–501. doi:http://dx.doi.org/10.1016/j.ejor.2006.07.009
3. Zhao Z, Fang J, Huang GQ, Zhang M (2016) Location management of cloud forklifts in finished product warehouse. Int J Intell Syst
4. Dai Q, Zhong R, Huang GQ, Qu T, Zhang T, Luo T (2012) Radio frequency identification-enabled real-time manufacturing execution system: a case study in an automotive part manufacturer. Int J Comput Integr Manuf 25(1):51–65
5. Ding B, Chen L, Chen D, Yuan H (2008) Application of RTLS in warehouse management based on RFID and Wi-Fi. In: 4th international conference on wireless communications, networking and mobile computing, 2008. WiCOM'08. IEEE, pp 1–5
6. Poon T, Choy KL, Chow HK, Lau HC, Chan FT, Ho K (2009) A RFID case-based logistics resource management system for managing order-picking operations in warehouses. Expert Syst Appl 36(4):8277–8301
7. Chow HK, Choy KL, Lee W, Lau K (2006) Design of a RFID case-based resource management system for warehouse operations. Expert Syst Appl 30(4):561–576
8. Choy K, Ho G, Lee C (2017) A RFID-based storage assignment system for enhancing the efficiency of order picking. J Intell Manuf 28(1):111–129
9. Li M, Gu S, Chen G, Zhu Z (2011) A RFID-based intelligent warehouse management system design and implementation. In: IEEE 8th international conference on e-Business Engineering (ICEBE), 2011. IEEE, pp 178–184
10. Mathworks (2014) Mixed-integer linear programming (MILP). https://se.mathworks.com/help/optim/ug/intlinprog.html

Improving the Decision-Making of Reverse Logistics Network Design Part II: An Improved Scenario-Based Solution Method and Numerical Experimentation

Hao Yu and Wei Deng Solvang

Abstract The study of the network design problems related to reverse supply chain and reverse logistics is of great interest for both academicians and practitioners due to its important role for a sustainable society. However, reverse logistics network design is a complex decision-making problem that involves several interactive factors and faces many uncertainties. Thus, in order to improve the reverse logistics network design, this paper proposes a new optimization model under stochastic environment and an improved solution method for network design of a multi-stage multi-product reveres supply chain. The study is presented in a series of two parts. Part I presents the relevant literature and formulates a stochastic mixed integer linear programming (MILP) for improving the decision-making of the reverse logistics network design. Part II improves the solution methods for the proposed stochastic programming and illustrates the application through a numerical experimentation.

Keywords Reverse logistics · Network design · Operational research
Optimization · Stochastic programming · MILP · Scenario-based solution

1 Introduction

Decision-making of a real world problem is hardly done with all relevant information available, but, in most cases, the decision has to be made even if some parameters cannot be accurately predicted or estimated at the time of decision-making [1]. For example, when a manufacturing facility is planned, the size of the facility is usually determined by the predication of future demand,

H. Yu (✉) · W. D. Solvang
Department of Industrial Engineering, UiT The Arctic University of Norway, Lodve Langesgate 2, 8514 Narvik, Norway
e-mail: hao.yu@uit.no

W. D. Solvang
e-mail: wei.d.solvang@uit.no

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_45

however, the prediction of the future is always wrong as discussed in many research works [2]. Thus, we have to make a proper decision with "the wrong predication of the future", and this will have a significant influence on the performance of the facility, say, an overoptimistic predication may lead to a waste of capacity or a high level of inventory, while a pessimistic estimation probably results in the incapability to fulfill the customer demands. With a deterministic model, this problem is usually tackled with a sensitivity analysis in order to determine the sensitivity of the optimal result to some key parameters. Nevertheless, the main shortcoming of this method is that sensitivity analysis can tell you to which degree the key parameters will affect the optimal result of the studied problem, but it cannot give you any suggestions on how to react to those changes [1]. Therefore, in reverse logistics network design, it is a much better way to treat the uncertainties in the modeling process with different techniques, i.e., fuzzy programming [3, 4], stochastic programming [5–7], robust optimization [8], etc.

In order to provide a better tool for reverse logistics network design under uncertainty, this research uses stochastic optimization in the modeling of the problem, and it also improves the solution method and presents some managerial implications. The first part of this research presents a MILP under stochastic environment for reverse logistics network design, and this paper focuses on the development of an improved multi-criteria scenario-based solution method and the numerical experimentation.

2 An Improved Scenario-Based Solution Method to the Stochastic Problem

The advantage of stochastic optimization is that it presents a proper model for the complex decision-making problem without the exact information of what is going to happen in the future [1]. Due to this reason, several possibilities and scenarios have to be accounted in a stochastic model in comparison with its deterministic form, which makes the model becoming more complex to solve. Many research works published in operational research and mathematical programming focus on the solution methods of stochastic optimization problems [9–12]. In this paper, based upon a recently published paper by Soleimani et al. [13], we develop an improved multi-criteria scenario-based solution method for the reverse logistics network design under stochastic environment.

Figure 1 presents the procedures of the solution method. As illustrated in the figure, the initial stage is to calculate the indicators for performance evaluation. There are three steps have to be conducted in this stage. First, the scenarios of the stochastic optimization problem are generated with respect to the change of uncertain parameters. Second, for each scenario generated, the problem becomes a MILP and can be resolved independently. The optimal reverse logistics network configurations calculated in each scenario are considered the candidates to the

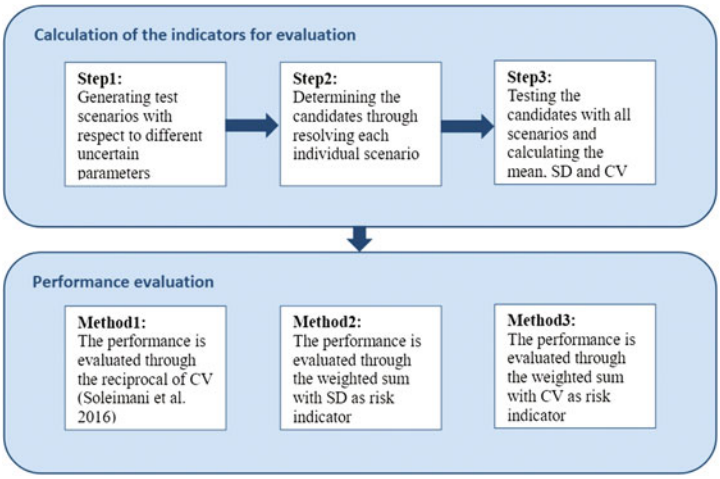


Fig. 1 The procedures of the multi-criteria scenario-based solution method with different performance evaluation

stochastic optimization problem. Third, all scenarios are tested for each candidate, and the indicators for evaluation including mean, standard deviation (SD) and coefficient of variation (CV) are calculated.

The performance of different candidates is compared with the reciprocal of CV in accordance with Soleimani et al.'s [13] method. The method aims to maximize the profits generated by the reverse logistics system while simultaneously minimize the risk of decision-making through the minimization of SD. SD is a very important measures of data dispersion in statistics and econometrics [14], say, how far in average the data dispersed from the mean. In reverse logistics network design, SD is used as the indictor for risk assessment, and a large SD implies a high risk on the realization of the expected profit. It is noteworthy that the risk or stability is not always considered an indicator for performance evaluation of the solutions to a stochastic problem, and the use of it heavily depends on the circumstance or problems modelled. For example, a venture capital may pursue a high return of investment (ROI), and the preference in this case is to have a high expectation of ROI and SD even if it leads to a high risk as well. However, the objective of reverse logistics network design is to have a stable profit generation with low risk, which gives long-term benefits to the players within the reverse supply chain. Thus, the evaluation with the reciprocal of CV is to maximize profit expectation while minimize the potential risk.

The evaluation method proposed by Soleimani et al. [13] has a significant shortcoming which may result in a sub-optimal solution dramatically affected by the risk or SD. For example, comparing with two candidates A (mean = 5000, SD = 1000) and B (mean = 1000, SD = 100), the reciprocal of CV of the two candidates are 5 and 10, respectively. Thus, the optimal solution is candidate B

according to the performance evaluation method. However, it is obvious that candidate A has a much higher profit expectation than candidate B even if the worst-case scenario happens. Candidate B is with very low potential risk, but the performance in profit generation is weak as well, so it is not a wise choice for the decision-maker. The reason is that when the mean and SD are composited, the relative importance of them is not accounted in the reciprocal of CV, so when there is a big difference in the mean, the optimal result calculated may be the one with both low risk and weak profitability. Besides, the evaluation method cannot be applied when the objective becomes a minimum function, e.g., min-cost, due to the simultaneous minimization of the mean and SD. In addition, there are also some challenges in the interpretation of the managerial implication as CV is used mainly for comparing the relative dispersion of data but not for decision-making of a stochastic problem [15].

$$Method_2 = \partial \frac{(M_{\max} - M_c)}{(M_{\max} - M_{\min})} + (1 - \partial) \frac{(SD_c - SD_{\min})}{(SD_{\max} - SD_{\min})} \tag{1}$$

$$Method_3 = \partial \frac{(M_{\max} - M_c)}{(M_{\max} - M_{\min})} + (1 - \partial) \frac{(CV_c - CV_{\min})}{(CV_{\max} - CV_{\min})} \tag{2}$$

In order to resolve the problems mentioned above, this paper further develop the multi-criteria scenario-based solution method with the help of a weighted sum function as illustrated in Eqs. (1) and (2). As shown in the equations, the mean and SD/CV are first normalized before they can be combined in the weighted sum. It is noteworthy that the potential risk is measured by both absolute indicator (SD) and relative indicator (CV), and the result will be compared in the numerical experimentation. Furthermore, the weighted sum formula can be easily modified in order to determine the optimal solution of a minimum objective function.

3 Numerical Experimentations

In this section, the proposed stochastic MILP model and improved scenario-based solution method are tested with a numerical experimentation. The problem includes 2 products, 15 customers, 8 candidates for collection and disassembly centers, 5 candidates for reuse/repair centers and 5 candidates for recycling centers. First, the parameters are given based upon uniform distribution. Then, three benchmark scenarios are defined as best-case (max-profit and max-product), deterministic (mean-profit and mean-product) and worst-case (min-profit and min-product), respectively. With the combination of the stochastic parameters, another 8 scenarios are generated in a logically reasonable way as illustrated in Table 1, meanwhile, the computational effectiveness is also considered in this process. For detailed introduction of scenario generation for a stochastic problem, the research works by King and Wallace [1] and Kaut and Wallace [16] can be referred.

Table 1 Generated scenarios for the numerical experimentation

Stochastic parameter	Scenario										
	Best	Basic	Worst	1	2	3	4	5	6	7	8
Amount of used products	Max	Mean	Min	L	L	L	L	H	H	H	H
Price/subsidy for reuse	Max	Mean	Min	L	L	H	H	L	L	H	H
Price/subsidy for recycling	Max	Mean	Min	L	H	L	H	L	H	L	H

Note H high, *L* low

The model is coded with Lingo programming and the optimal solution of each scenario is first calculated as a MILP problem. Figure 2 shows the profit generated from the reverse logistics system in each candidate solution and the mean value. It is clearly that there are a fair allocation of the tested scenarios around the mean, this represents a large variety of potential scenarios may happen in future under market fluctuation with both optimistic and pessimistic expectation. Even if more scenarios can be generated, but as many argues [5, 7], this dramatically increases the complexity of the problem, but the benefit yielded is extremely limited. Thus, the scenarios generated in this numerical experiment is in a rational and efficient way.

In a stochastic optimization problem, the objective is to determine the optimal solution through all the scenarios. As discussed by King and Wallace [1], there are two characteristics of the optimal solution of a stochastic problem, namely, robustness and flexibility. In some situations, the optimal solution should be robust to withstand all the possible situations in future without many changes, e.g., a bus schedule, while in many other circumstances, the focus of a stochastic programming is to be flexible and easy to adapt to the changes and fluctuations, e.g., a schedule or plan of a fast delivery company. In the reverse logistics network design, both robustness and flexibility are focused in the decision-making, say, the long-time strategic decisions (locations of different facilities) should be robust and remain unchanged for a long period, while the short-term tactical decisions (distribution of used products and transportation strategies) should be flexible and

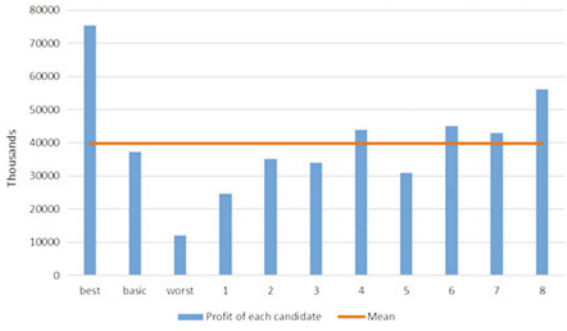


Fig. 2 The profit of each candidate and the mean

updated periodically in order to maximize the profit generated under market fluctuation. Therefore, the stochastic MILP model formulated in this paper is two-stage in nature. The first stage variables are facility location decisions, and the second stage variables are the decisions related to the distribution and transportation of used products.

In the second step of the solution method, the candidates are tested through all scenarios. The first stage decisions will not change, while the second stage decisions are flexible with the changing price and generated amount of used products in order to maximize the economic performance of the reverse logistics system. The comparison of the profit expectation (mean) and level of risk (SD) is given in Fig. 3. As shown in the figure, the expected profit is higher with the candidate solutions calculated under high amount of used products generated, while the level of risk is better maintained with the candidates calculated under low generation of used products.

In the testing phase, we observed infeasibilities when some candidate reverse logistics network configurations are implemented in the scenarios with high generation of used products. The reason for the infeasibility is the conflict between service constraint and capacity constraint, say, the facility capacity is not large enough to handle all the used products generated. In order to solve this problem, the model needs to be relaxed and a compromise has to be made based on the decision-maker's objective, either to provide a reduced service or to have more investment for increasing the capacity of reverse logistics system. In this paper, the relaxation of the capacity constraint is made in some scenarios in order to generate feasible solutions, and the relaxation is made in such a way that the changes to the original network configuration are kept at the minimum level.

The performance of the candidates is measured with three different methods, and the result is given in Table 2. For implementing the weighted sum methods, several combinations of weights are tested and $\partial = 0.7$ is selected for methods 2 and 3. As illustrated in the table, the optimal results from different evaluation methods vary significantly. The optimal solutions evaluated by methods 1, 2 and 3 are candidates 1, 7 and 5, respectively.

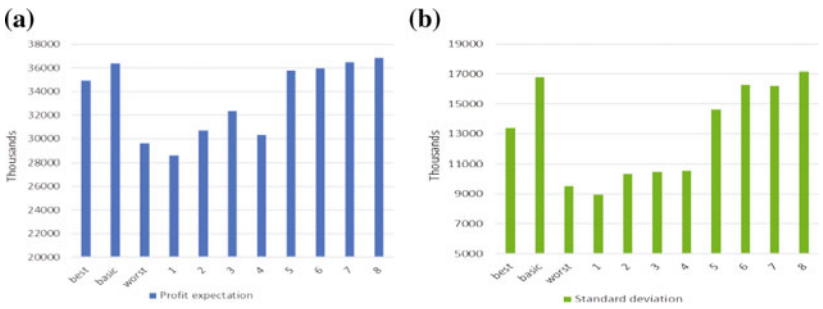


Fig. 3 The performance of candidates through all scenarios: **a** profit expectation; **b** standard deviation

Table 2 Comparison of the performance evaluation with the three methods

Scenarios	Evaluation method		
	Method1	Method2	Method3
Best	2.6099	0.3257	0.3016
Basic	2.1698	0.3253	0.3304
Worst	3.1043	0.6347	0.6309
1	*3.1933*	0.7000	0.7000
2	2.9605	0.5746	0.5715
3	3.0837	0.4397	0.4057
4	2.8748	0.6109	0.6208
5	2.4456	0.3002	*0.2815*
6	2.2069	0.3465	0.3543
7	2.2494	*0.2988*	0.2923
8	2.1489	0.3000	0.3000

Figure 4 presents the comparison of the mean and profit expectation in each scenario of the three candidates. As shown in the figure, the overall profit expectation of candidate 1 is much lower than that of candidates 7 and 5, and the overall profit expectation of candidate 7 is slightly higher than that of candidate 5. When the amount of used products generated is low, candidate 1 has a slightly better performance than the others, but when the generation of used products is high, the economic performance of candidates 7 and 5 are much better. Thus, it is obvious that the optimal solution calculated by the reciprocal of CV is with lower level of risk and lower economic performance. Compared with candidate 1, even if the level of risk is higher, both candidates 7 and 5 have much better economic performance particularly when the generation of used products is high. To distinguish with the two candidates, candidate 7 has a slightly higher profit expectation of the reverse logistics system, while candidate 5 has a slightly better level of risk (slightly better performance in low generation scenarios).

From the discussion above, it is clearly that the optimal solution calculated by the original method is a weak performance one even if the level of risk is low. Both of the other methods with weighted sum can generate better solutions compared

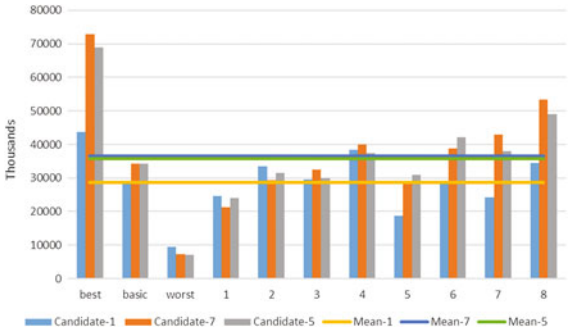


Fig. 4 Comparison of the mean and profit expectation in test scenarios of the optimal solutions calculated with the three different method

with the original method, and the optimal solution may be selected based upon the decision-maker's preference.

4 Conclusion

In this paper, we developed a new MILP model under stochastic environmental for designing a multi-product multi-stage reverse logistics network, and an improved multi-criteria scenario-based solution method was also proposed for resolving the stochastic optimization problem. The objective of the proposed model is to maximize the overall profit generated in the reverse logistics system under market fluctuation, and the solution method in a recently published research work is improved so that both profit expectation and potential risk are taken into consideration in the decision-making. A numerical experimentation is conducted, and the application of the model and the effectiveness of the improved solution method are explicitly presented.

For future improvement of the research, the consideration of environmental and social sustainability should be taken into account in the reverse logistics network design [4, 17]. The simultaneous consideration of several conflict objectives under uncertain environment may significantly increase the computational complexity, so the development of advanced solution algorithms is also suggested as a promising direction for the future research.

References

1. King AJ, Wallace SW (2012) Modeling with stochastic programming. Springer Science & Business Media
2. Chopra S, Meindl P (2007) Supply chain management. Strategy, planning & operation. Das summa summarum des Manage 265–275
3. Chu LK, Shi Y, Lin S, Sculli D, Ni J (2010) Fuzzy chance-constrained programming model for a multi-echelon reverse logistics network for household appliances. Oper Res Soc 61 (4):551–560
4. Govindan K, Paam P, Abtahi A-R (2016) A fuzzy multi-objective optimization model for sustainable reverse logistics network design. Ecol Ind 67:753–768
5. El-Sayed M, Afia N, El-Kharbotly A (2010) A stochastic model for forward–reverse logistics network design under risk. Comput Ind Eng 58(3):423–431
6. Ramezani M, Bashiri M, Tavakkoli-Moghaddam R (2013) A new multi-objective stochastic model for a forward/reverse logistic network design with responsiveness and quality level. Appl Math Model 37(1–2):328–344
7. Pishvaee MS, Jolai F, Razmi J (2009) A stochastic optimization model for integrated forward/reverse logistics network design. J Manufact Syst 28(4):107–114
8. Talaei M, Moghaddam BF, Pishvaee MS, Bozorgi-Amiri A, Gholamnejad S (2016) A robust fuzzy optimization model for carbon-efficient closed-loop supply chain network design problem: a numerical illustration in electronics industry. J Clean Prod 113:662–673
9. Wallace SW, Ziemba WT (2005) Applications of stochastic programming: SIAM

10. Kall P, Wallace SW, Kall P (1994) Stochastic programming. Springer, Berlin
11. Higle JL, Wallace SW (2003) Sensitivity analysis and uncertainty in linear programming. Interfaces 33(4):53–60
12. Birge JR, Louveaux F (2011) Introduction to stochastic programming. Springer Science & Business Media
13. Soleimani H, Seyyed-Esfahani M, Shirazi MA (2016) A new multi-criteria scenario-based solution approach for stochastic forward/reverse supply chain network design. Ann Oper Res 242(2):399–421
14. Washington SP, Karlaftis MG, Mannering F (2010) Statistical and econometric methods for transportation data analysis. CRC press
15. Yu H, Solvang W (2016) A stochastic programming approach with improved multi-criteria scenario-based solution method for sustainable reverse logistics design of waste electrical and electronic equipment (WEEE). Sustainability 8(12):1331
16. Kaut M, Wallace SW (2003) Evaluation of scenario-generation methods for stochastic programming
17. Yu H, Solvang WD (2016) A general reverse logistics network design model for product reuse and recycling with environmental considerations. Int J Adv Manufact Technol 87 (9–12):2693–2711

Improving the Decision-Making of Reverse Logistics Network Design Part I: A MILP Model Under Stochastic Environment

Hao Yu and Wei Deng Solvang

Abstract The study of the network design problems related to reverse supply chain and reverse logistics is of great interest for both academicians and practitioners due to its important role for a sustainable society. However, reverse logistics network design is a complex decision-making problem that involves several interactive factors and faces many uncertainties. Thus, in order to improve the reverse logistics network design, this paper proposes a new optimization model under stochastic environment and an improved solution method for network design of a multi-stage multi-product reveres supply chain. The study is presented in a series of two parts. Part I presents the relevant literature and formulates a stochastic mixed integer linear programming (MILP) for improving the decision-making of the reverse logistics network design. Part II improves the solution method for the proposed stochastic programming and illustrates the application through a numerical experimentation.

Keywords Reverse logistics · Network design · Operational research
Optimization · Stochastic programming · MILP · Scenario-based solution

1 Introduction

Supply chain management is a well-developed concept, and it has been extensively focused due to its significant role in determining the success of a company. As argued by Chopra and Meindl [1], in today's dynamic and competitive marketplace, the competition is no longer a company versus company game, but it is a competition between different supply chains. Thus, the design of an effective supply

H. Yu (✉) · W. D. Solvang
Department of Industrial Engineering, UiT The Arctic University of Norway, Lodve Langesgate 2, 8514 Narvik, Norway
e-mail: hao.yu@uit.no

W. D. Solvang
e-mail: wei.d.solvang@uit.no

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_46

chain is of great importance for a company to survive in the market and gain profits [2]. Traditionally, the focus of supply chain management is to manage the material, cash and information flow from the raw material suppliers, via manufacturers, distributors and/or retailers, towards the end customers in order to generate profits through the fulfilment of customer demands [1, 3]. The primary objective of such supply chain network design problems becomes therefore profit- or cost-focused [4]. However, with the increased emphasis on circular economy and sustainable development in recent years, the management of environmental impact of supply chain activities has been gained significant attentions [5], and new concepts, i.e., green supply chain, reverse logistics, etc., have been introduced and widely accepted by the public as well as that in the research world.

Reverse logistics is one of the most important means for improving the environmental performance of a supply chain through reducing the waste and recovering value from used products [6, 7]. During the latest two decades, the study of the network design problems related to reverse logistics and reverse supply chain has been of significant interest for both academicians and practitioners [8]. Critical literature reviews are given by different researchers to summarize the current development and suggest new directions for future research [9–12]. In this paper, some of the latest research works in reverse logistics network design are reviewed.

Li et al. [13] develop a hybrid artificial bee colony algorithm for the network design problem of reverse logistics system, which minimizes the overall system operating costs. Considering the cost-minimization of an integrated forward/reverse logistics system, Tsao et al. [14] formulate a non-linear mathematical programming incorporating with a continuous approximation method for reverse logistics network design. Zandieh and Chensebli [15] investigate a water flow like algorithm for minimizing the costs related to the establishment and operation of a reverse logistics system. Taking into account of the cross-docking operations in reverse logistics design, Kheirkhah and Rezaei [16] formulate a mixed integer linear programming (MILP) in order to minimize the overall costs.

Reverse logistics network design usually involves several objectives. Yu and Solvang [7] propose a bi-objective optimization model with a normalized weighed sum method for sustainable reverse logistics network design, which considers the balance of costs and carbon emissions. Yilmaz et al. [17] develop a bi-objective MILP model for optimal route planning problem of a hazardous waste management system, and it aims at minimizing both costs and environmental risks. Govindan et al. [18] investigate a fuzzy multi-objective MILP model simultaneously balancing the economic, environmental and social sustainability in reverse logistics design. Yu and Solvang [19] formulate a multi-objective optimization for sustainable design of a municipal solid waste (MSW) management system.

Decision-making under inexact input information is another well-focused topic in reverse logistics network design. Keshavarz Ghorabaee et al. [4] investigate a two stage decision-making system for designing an integrated forward/reverse logistics system under uncertainties. Soleimani and Govindan [20] develop a stochastic optimization model for reverse logistics network design considering the minimization of risk in decision-making. Talaei et al. [21] propose a robust fuzzy

optimization model to account both costs and carbon emissions of reverse logistics design under uncertainties.

In the real world, decision-making is never done with all necessary information available [22]. Therefore, it is of significance to develop advanced tool for decision-making of reverse logistics network design with inexact information. This paper investigates a new MILP model under stochastic environment for reverse logistics network design, and the generation of used products and price for reused and recycled products are formulated as stochastic parameters. The study is presented in a series of two parts. Part I formulates the mathematical model. Part II improves the solution method and illustrates the application through a numerical experimentation.

2 Problem and Modeling

This paper focuses on an open-loop supply chain network structure as presented in Yu et al. [23]. Figure 1 illustrates that the reverse logistics system includes five types of nodes: end customer, collection and dissembling center, reuse/repair center, recycling center and waste disposal plant, respectively. As can be seen in the figure, the flow of used products in the open-loop reverse logistics system starts from the end customers, via intermediate nodes for collection, disassembly, repair and recycling, towards different primary and secondary markets for value recover and landfills for proper disposal.

It is a prerequisite that the relevant parameters are known or can be estimated for decision-making. Moreover, in order to formulate the problem, the notations of the indices, sets, parameters, decision variables are first given in Table 1.

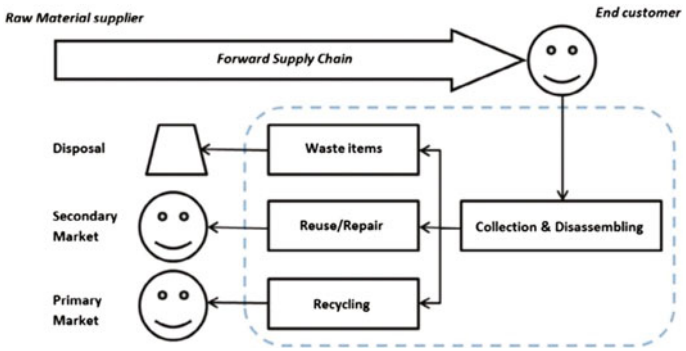


Fig. 1 The structure of reverse logistics network [23]

Table 1 Notations of indices, sets parameters and decision variables used in the formulation of the mathematical model

Sets and indices	
S	Set of scenarios, $s \in S$
P	Set of products, $p \in P$
M	Set of markets/customers, $m \in M$
C	Set of potential locations for collection centers, $c \in C$
J	Set of potential locations for repair centers, $j \in J$
K	Set of potential locations for recycling centers, $k \in K$
L	Set of disposal centers, $l \in L$
Parameters	
F_c, F_j, F_k	Fixed costs for collection center c, repair center j and recycling center k
V_{cp}, V_{jp}, V_{kp}	Variable costs at collection center c, repair center j and recycling center k
V_{lp}	Treatment costs at established disposal center l
T_{mcp}	Unit costs for first level transportation of product p between customer m and collection center c
T_{cjp}, T_{ckp}, T_{clp}	Unit costs for second level transportation of product p between collection center c and repair center j, between collection center c and recycling center k, and between collection center c and disposal center l
W_{mps}	Generation of used products p at customer m in scenario s
CR_{pj}, CR_{pk}	Reusable and recyclable fractions of product p
B_{cp}, B_{jp}, B_{kp}	Planned capacity of collection center c, repair center j and recycling center k for product p
SP_{pjs}, SD_{pjs}	Profit and government subsidy for repairing one unit of product p
SP_{pks}, SD_{pks}	Profit and government subsidy for recycling one unit of product p
Decision variables	
ϑ_{cs}, ϑ_{js}, ϑ_{ks}	First stage decision variables determining if the potential locations for collection center, repair center and recycling center are opened in scenario s
AM_{mcps}	Quantity of first level transportation of used products p between customer m and collection center c in scenario s
AM_{cjps}, AM_{ckps}, AM_{clps}	Quantity of first level transportation of used products p between collection center c and repair center j, between collection center c and recycling center k, and between collection center c and disposal center l in scenario s

2.1 Constraints

The primary function of reverse logistics is to properly reuse, recycle, treat and dispose the used products, so Eq. (1) guarantees the used products from all customers/markets are collected.

$$\sum_{c \in C} AM_{mcps} = W_{mps}, \quad \forall m \in M, \ \ p \in P, \ \ s \in S \tag{1}$$

The equilibrium constraints are formulated in Eqs. (2)–(4). Equation (2) ensures the input amount of used products from all customers equals to the output amounts for repair, recycling and disposal. Constraints (3) and (4) are to guarantee the amounts of used products sent for reuse and recycling are less than or equal to the respective fractions, which means the components distributed to different facilities must be complied with the reuse/recycling technologies implemented.

$$\sum_{m\in M} AM_{mcps} = \sum_{j\in J} AM_{cjps} + \sum_{k\in K} AM_{ckps} + \sum_{l\in L} AM_{clps}, \quad \forall c \in C,\ p \in P,\ s \in S \tag{2}$$

$$CR_{pj} \sum_{m\in M} AM_{mcps} \geq \sum_{j\in J} AM_{cjps}, \quad \forall c \in C,\ p \in P,\ s \in S \tag{3}$$

$$CR_{pk} \sum_{m\in M} AM_{mcps} \geq \sum_{k\in K} AM_{ckps}, \quad \forall c \in C,\ p \in P,\ s \in S \tag{4}$$

Capacity limitation constraints are formulated as follows. Equations (5)–(8) guarantee the amount of different types of used products treated at repair center, recycling center and disposal center are less than or equal to the respective capacities. The influence of product type is negligible when they are sent to landfills.

$$\sum_{m\in M} AM_{mcps} \leq \vartheta_{cs} B_{cp}, \quad \forall c \in C,\ p \in P,\ s \in S \tag{5}$$

$$\sum_{c\in C} AM_{cjps} \leq \vartheta_{js} B_{jp}, \quad \forall j \in J,\ p \in P,\ s \in S \tag{6}$$

$$\sum_{c\in C} AM_{ckps} \leq \vartheta_{ks} B_{kp}, \quad \forall k \in K,\ p \in P,\ s \in S \tag{7}$$

$$\sum_{c\in C} \sum_{p\in P} AM_{clps} \leq B_l, \quad \forall l \in L,\ s \in S \tag{8}$$

Equations (9) and (10) are constraints for binary and non-negative decision variables.

$$\vartheta_{cs}, \vartheta_{js}, \vartheta_{ks} \in \{0, 1\}, \quad \forall c \in C,\ j \in J,\ k \in K,\ s \in S \tag{9}$$

$$AM_{mcps}, AM_{cjps}, AM_{ckps}, AM_{clps}, \quad \forall m \in M,\ c \in C,\ j \in J,\ k \in K,\ l \in L,\ p \in P,\ s \in S \tag{10}$$

2.2 Objective Function

The objective of the reverse logistics network design is to maximize the total profits generated from the reuse and recycling of used products. In order to improve the enthusiasm of companies to take part in the reverse logistics, the influence of government subsidy is taken into consideration in this model, so the total revenue of the reverse logistics system is comprised of both sales revenue and subsidy.

$$Profit = Revenue - Cost,\ s \in S \tag{11}$$

$$\begin{aligned} Revenue = & \sum_{c\in C}\sum_{p\in P}\sum_{j\in J}(SP_{pjs} + SD_{pjs})AM_{cjps} \\ & + \sum_{c\in C}\sum_{p\in P}\sum_{k\in K}(SP_{pks} + SD_{pks})AM_{ckps},\ s \in S \end{aligned} \tag{12}$$

$$\begin{aligned} Cost = & \sum_{c\in C} F_c\vartheta_{cs} + \sum_{j\in J} F_j\vartheta_{js} + \sum_{k\in K} F_k\vartheta_{ks} \\ & + \sum_{m\in M}\sum_{c\in C}\sum_{p\in P} V_{cp}AM_{mcps} + \sum_{c\in C}\sum_{j\in J}\sum_{p\in P} V_{jp}AM_{cjps} \\ & + \sum_{c\in C}\sum_{k\in K}\sum_{p\in P} V_{kp}AM_{ckps} + \sum_{c\in C}\sum_{d\in D}\sum_{p\in P} V_{dp}AM_{clps} \\ & + \sum_{m\in M}\sum_{c\in C}\sum_{p\in P} T_{mcp}AM_{mcps} + \sum_{c\in C}\sum_{j\in J}\sum_{p\in P} T_{cjp}AM_{cjps} \\ & + \sum_{c\in C}\sum_{k\in K}\sum_{p\in P} T_{ckp}AM_{ckps} + \sum_{c\in C}\sum_{d\in D}\sum_{p\in P} T_{clp}AM_{clps},\ s \in S \end{aligned} \tag{13}$$

The objective function of the proposed model is given in Eqs. (11)–(13). Equation (11) shows the overall profit of the reverse logistics system equals to the total revenue minus total costs. Equation (12) calculates the total revenue from the product recovery and government subsidy. Equation (13) calculates the total costs for establishing and operating the reverse logistics system including facility opening costs, variable costs and transportation costs. The optimal solution to the MILP model under stochastic environment is not the optimal solution to a single scenario but the one with the best profit expectation through all the scenarios.

3 Summary

With the increasing focus on the circular economy and sustainable development all over the world, the research on reverse logistics has become a highly focused topic. In this paper, recently published research works in reverse logistics network design are first reviewed and then a new MILP model under stochastic environment is

formulated in order to improve the decision-making of reverse logistics network design under inexact input information. In this model, the generation of used products at different markets/customers and the price for reused and recycled products are considered as stochastic parameters. This research is presented in a series of two parts. In the second part, the proposed MILP model under stochastic environment is resolved by an improved multi-criteria scenario-based solution method, and the application of the model is illustrated through a numerical experimentation.

References

1. Chopra S, Meindl P (2007) Supply chain management. Strategy, planning and operation. Das summa summarum des Manage 265–275
2. Moncayo–Martínez LA, Mastrocinque E (2016) A multi-objective intelligent water drop algorithm to minimise cost of goods sold and time to market in logistics networks. Expert Syst Appl 64:455–466
3. Hugos MH (2011) Essentials of supply chain management. Wiley, New York
4. Keshavarz Ghorabaee M, Amiri M, Olfat L, Khatami Firouzabadi SA (2017) Designing a multi-product multi-period supply chain network with reverse logistics and multiple objectives under uncertainty. Technol Econ Dev Econ 23(3):520–548
5. Beamon BM (1999) Designing the green supply chain. Logistics Inf Manage 12(4):332–342
6. Diabat A, Kannan D, Kaliyan M, Svetinovic D (2013) An optimization model for product returns using genetic algorithms and artificial immune system. Resour Conserv Recycl 74:156–169
7. Yu H, Solvang WD (2016) A general reverse logistics network design model for product reuse and recycling with environmental considerations. Int J Adv Manufact Technol 87(9–12):2693–2711
8. Shekarian E, Olugu EU, Abdul-Rashid SH, Bottani E (2016) A fuzzy reverse logistics inventory system integrating economic order/production quantity models. Int J Fuzzy Syst 18 (6):1141–1161
9. Agrawal S, Singh RK, Murtaza Q (2015) A literature review and perspectives in reverse logistics. Resour Conserv Recycl 97:76–92
10. Govindan K, Soleimani H, Kannan D (2015) Reverse logistics and closed-loop supply chain: a comprehensive review to explore the future. Eur J Oper Res 240(3):603–626
11. Pokharel S, Mutha A (2009) Perspectives in reverse logistics: a review. Resour Conserv Recycl 53(4):175–182
12. Mahaboob Sheriff K, Gunasekaran A, Nachiappan S (2012) Reverse logistics network design: a review on strategic perspective. Int J Logistics Syst Manage 12(2):171–194
13. Li J-q, Wang J-d, Pan Q-k, Duan P-y, Sang H-y, Gao K-z et al (2017) A hybrid artificial bee colony for optimizing a reverse logistics network system. Soft Comput 1–18
14. Tsao Y-C, Linh V-T, Lu J-C, Yu V (2017) A supply chain network with product remanufacturing and carbon emission considerations: a two-phase design. J Intell Manufact 1–13
15. Zandieh M, Chensebli A (2016) Reverse logistics network design: a water flow-like algorithm approach. OPSEARCH 53(4):667–692
16. Kheirkhah A, Rezaei S (2016) Using cross-docking operations in a reverse logistics network design: a new approach. Prod Eng Res Dev 10(2):175–184

17. Yilmaz O, Kara BY, Yetis U (2016) Hazardous waste management system design under population and environmental impact considerations. J Environ Manage. https://doi.org/10.1016/j.jenvman.2016.06.015
18. Govindan K, Paam P, Abtahi A-R (2016) A fuzzy multi-objective optimization model for sustainable reverse logistics network design. Ecol Ind 67:753–768
19. Yu H, Solvang WD (2017) A multi-objective location-allocation optimization for sustainable management of municipal solid waste. Environ Syst Decisions 37(3): 289-308
20. Soleimani H, Govindan K (2014) Reverse logistics network design and planning utilizing conditional value at risk. Eur J Oper Res 237(2):487–497
21. Talaei M, Moghaddam BF, Pishvaee MS, Bozorgi-Amiri A, Gholamnejad S (2016) A robust fuzzy optimization model for carbon-efficient closed-loop supply chain network design problem: a numerical illustration in electronics industry. J Clean Prod 113:662–673
22. King AJ, Wallace SW (2012) Modeling with stochastic programming. Springer Science & Business Media
23. Yu H, Solvang W, Solvang B (2016) A multi-objective mathematical programming for sustainable reverse logistics network design. Part I: model formulation. WIT Trans Eng Sci 113:287–295

Predictive Maintenance for Synchronizing Maintenance Planning with Production

Harald Rødseth, Per Schjølberg, Markus Wabner and Uwe Frieß

Abstract With the immense pressure to sustain competitive in European manufacturing, the strategy of digitalizing in this industry sector is indeed necessary. With the onset of new ICT technology and big data capabilities, the physical asset and data computation is integrated in manufacturing through Cyber Physical Systems (CPS). This strategy also denoted as Industry 4.0 will also improve the maintenance function significantly in manufacturing. In particular, maintenance planning will be more synchronized in production scheduling. The aim of this article is to develop an integrated planning (IPL) approach that synchronizes production and maintenance planning with predictive maintenance capability. The result in this article is based on a case study and simulation of manufacturing equipment. In particular, application of key performance indicators (KPIs) is shown to be essential when running the synchronizing mechanism in IPL. The scientific application of the case study is alignment of the IPL theory and new approach in maintenance planning. Furthermore, the application to practice is improved maintenance planning in IPL that increases a reliable plant capacity. It is concluded that the IPL approach should be considered to be a generic platform for manufacturing industry that should be demonstrated further in other manufacturing branches in Europe.

Keywords Predictive maintenance · Integrated planning · Industry 4.0
Key performance indicators

H. Rødseth (✉) · P. Schjølberg
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology (NTNU), NO 7491, Trondheim, Norway
e-mail: Harald.Rodseth@ntnu.no

P. Schjølberg
e-mail: Per.Schjolberg@ntnu.no

M. Wabner · U. Frieß
Department of Machine Tools, Fraunhofer IWU, 09126 Chemnitz, Germany
e-mail: markus.wabner@iwu.fraunhofer.de

U. Frieß
e-mail: uwe.friess@iwu.fraunhofer.de

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_47

1 Introduction

It is an important need in European manufacturing to sustain competitive supported by information and communication technology [8]. With the internet age, the industry is now offered a new concept to process information and control the manufacturing equipment with Internet of Things (IoT) and big data analytics. In sum, this concept is denoted as Industry 4.0 [9].

With the possibilities offered in Industry 4.0, maintenance planning in predictive maintenance can benefit of being more synchronized with other disciplines such as production. This is also supported by the European Commission where methods that schedule maintenance activities together with production activities are developed and demonstrated [5]. These methods are also known to be a part of the integrated planning (IPL) approach [14]. In particular, a structured approach for data driven predictive maintenance has been proposed [16]. It remains to investigate how predictive maintenance algorithms can support IPL.

The aim in this article is to develop and demonstrate a novel IPL approach that synchronizes production and maintenance planning with predictive maintenance capability.

The further structure of this article is as follows: Sect. 2 presents the state of the art within digitalization of maintenance, IPL and predictive maintenance. The IPL approach is presented in Sect. 3 based on a case study. Finally, concluding remarks are given in Sect. 4.

2 Literature

2.1 Digitalizing Maintenance

Based on Industry 4.0, many future trends are related to maintenance. In Norway, two important technology driven trends have been pointed out to be relevant for maintenance [13]:

- *Digitalizing the Industry*: Smart application of sensor technology can both reduce the down time for machines and improve the efficiency of maintenance.
- *Virtual Reality*: Application of augmented reality as projections on glasses can support maintenance engineers in visualizing the effect of conducting maintenance.

Also, the German standardization roadmap for Industry 4.0 endorses maintenance and presents *smart maintenance* to be an enabler of Industry 4.0 [4]. It is even stated that without systematic development of maintenance into smart maintenance, the successful implementation of Industry 4.0 will be put at risk.

In particular, predictive maintenance has been pointed out to be an important element in digitalizing maintenance with different application within remote maintenance [11]. It has also been estimated that predictive maintenance can reduce the machine downtime by 30–50% and extend the lifetime of the machine by 20–40% [12].

2.2 *Remaining Useful Life Algorithm in Predictive Maintenance*

Model based approaches in general try to reproduce certain aspects of real-life physical assets like machine tools. In case of condition monitoring the wear-causing mechanism has to be reproduced by the underlying model. This inherently leads to a prediction functionality of model based approaches [2, 3, 6, 7]. Therefore model-based approaches are not only usable for in-service monitoring and RUL prediction, but also as estimation of the life expectancy of components [3]. Figure 1 shows characteristics of model-based approaches.

However the challenge of model-based approaches comes from the indirect linkage of wear-causing measuring values, e.g. loads over time, to the life-expectancy of components. The failure modes are typically stochastically determined and influenced by various parameters up until tolerances of ball elements or the actual state of the lubricants. This inherently reduces the accuracy of empirical models [2, 7].

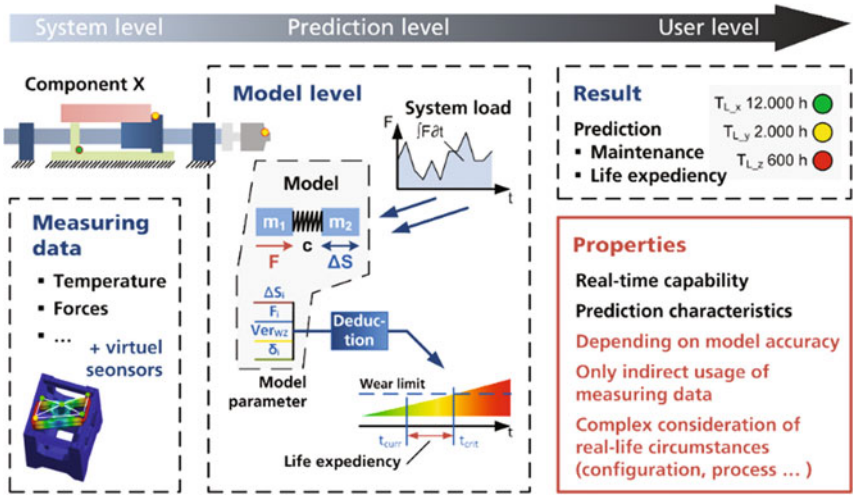


Fig. 1 Characteristics of model-based RUL methods

2.3 Predictive Maintenance and Maintenance Planning

The European standard EN 13306 [1] defines predictive maintenance as "*Condition based maintenance carried out following a forecast derived from repeated analysis or known characteristics and evaluation of the significant parameters of the degradation of the item*." This standard also classifies condition based maintenance as preventive maintenance where maintenance is carried out at predetermined intervals or other criteria intended to reduce the probability of failure. In IPL a reliability based approach for maintenance optimization has been proposed that establish the maintenance intervals [15]. The heuristic for indirect maintenance grouping has been presented in [15]. By following this heuristic, maintenance activities are scheduled with specific maintenance intervals where maintenance costs and reliability data are considered. For the reliability data, Weibull distribution is assumed to model degradation of the system. Maintenance costs comprise preventive maintenance costs, corrective maintenance costs, and set-up costs. Further research for this model is to include predictions of remaining useful life (RUL). This estimation will improve the decision support e.g. during the production planning. In particular it will be of interest to study how model based approaches will build up a work order system in maintenance planning where estimation of RUL will provide digital support in maintenance notifications. This will then require more investigation developing a digitalized predictive maintenance architecture. Currently, the 5C architecture for implementation of Cyber-Physical System proposed by Lee [10] is of highly relevance and has been tested in different industry branches such as the aluminum production [17].

3 IPL Approach

The proposed IPL approach is outlined in Fig. 2. Further in this chapter each step is elaborated. Also, the model of RUL is based on a case study from wear of bearings.

3.1 Step 1: Initial Maintenance Plan

Maintenance planning is essential in preventive maintenance where a detailed maintenance plan is created. The maintenance plan is defined by EN 13306 [1] as

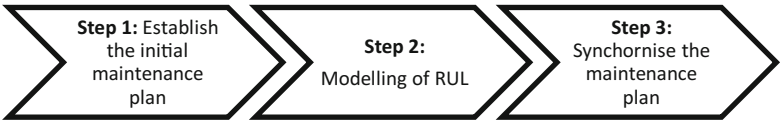


Fig. 2 IPL approach of maintenance synchronisation

"*Structured and documented set of tasks that include the activities, procedures, resources and the time scale required to carry out maintenance*." To establish an initial maintenance plan for IPL, indirect grouping is proposed in this phase, see Fig. 3. For indirect grouping there are only opportunities for preventive maintenance each T time units. For each component i, it is possible to perform preventive maintenance every $k_i T$ time unit. k_i is an integer.

3.2 Step 2: Modelling of RUL

The empirical formula regarding balls screws and linear guideways with rolling ball elements is well-established [6]:

$$L_{10}^{0} = \left(\left[\left(\frac{C_a \times T_{0.1}}{\int_{T=0}^{T_{0.1}} F \partial t \times k_{1\ldots n}} \right)^3 \times 10^6 \right]^{-1} + \left[\left(\frac{C_a \times [T_{0.2} - T_{0.1}]}{\int_{T_{0.1}}^{T_{0.2}} F \partial t \times k_{1\ldots n}} \right)^3 \times 10^6 \right]^{-1} + \cdots + \left[\left(\frac{C_a \times [T_1 - T_{i-1}]}{\int_{T_{i-1}}^{T_1} F \partial t \times k_{1\ldots n}} \right)^3 \times 10^6 \right]^{-1} \times i \right) \tag{1}$$

where

L_{10}^{0} Overall life expectancy with 10% failure probability [distance, e.g. km/rotations]
I sector i
C_a Dynamic parameter (supplier value based on experimental test)
$T_{0.i}$ Period 0, subsector i
T_1 Overall first period
F Load in sector i
k_n Correction parameters

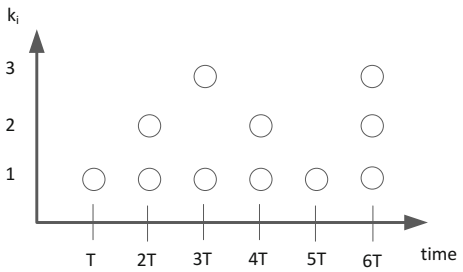


Fig. 3 Initial maintenance plan, based on [15]

It directly predicts the lifetime of a component based on wear-causing load over time. By using real-life data about historical loads and speeds it offers the possibility to directly calculate the already used percentage of the "wear stock" and calculate RUL. Furthermore, RUL can be evaluated virtually before performing certain manufacturing processes on a machine tool.

Based on the empirical formula an algorithm, a data acquisition methodology and a graphic user interface were developed to directly interact and calculate the RUL based on real-life data, see Fig. 4.

Historic load data of components are stored and can be used to calculate RUL. In addition, any artificial load-cycle or combination of several load-cycles can be used as well. Also an average lifetime of a given component can be estimated for a given process—translated to a concrete load on the component level—and/or compare how fast a given process will eat up the wear stock of different machines available for production.

3.3 Step 3: Synchronizing Maintenance Plan

When RUL has been modelled, this can be applied for synchronizing the maintenance and production plan. Figure 5 illustrates an example of how the maintenance synchronization is conducted.

At 4T there is an overlap between the maintenance and production plan, and it is of interest to evaluate the possibilities of restructuring the maintenance plan. When considering the analysis of the component at t_0, two possible scenarios can be identified; RUL1 and RUL2. If RUL1 is the most likely scenario, it is not possible

Fig. 4 Graphic user interface (GUI)

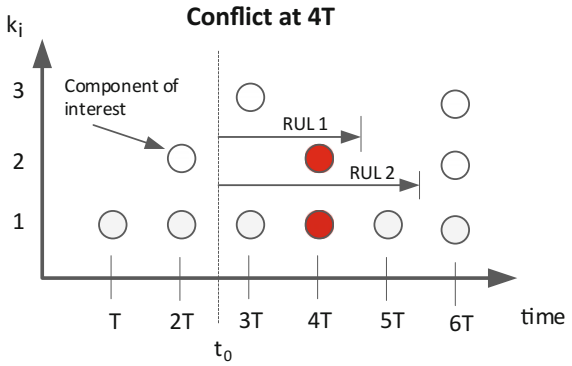


Fig. 5 Maintenance-production synchronization

to postpone the maintenance activity to 5T and it should instead be conducted at 3T. For RUL2 it is possible to postpone the maintenance activity to 5T.

4 Concluding Remarks

The aim in this article was to develop and demonstrate an IPL approach. With methodology from maintenance grouping and predictive maintenance, and IPL approach with three steps was proposed.

It is concluded that this IPL approach is a novel predictive maintenance concept combining maintenance grouping with a model based prediction of RUL.

Further research will require more detailed and integrated modelling of maintenance scheduling and prediction of RUL. For example, based on historical estimation of RUL it can be possible to re-adjust the initial maintenance plan.

Another topic that is important to consider in future research is the balance between accuracy in estimating RUL and the future time window in maintenance planning. For the maintenance planner it is important to have a future time window in order to reschedule when there is an overlap between the maintenance and production plan. This will enable the maintenance planner to reschedule the maintenance activity either earlier or later than originally planned. However, when a larger maintenance window is required it can be more difficult to predict the RUL since the future load characteristics of the bearing can fluctuate.

Finally, it is expected that more testing in industry is needed for the proposed IPL approach including several industry branches.

References

1. CEN (2010) EN 13306: Maintenance—maintenance terminology
2. Denkana B, Bluemel P, Kroening S, Roebbing J (2011) Condition based maintenance planning of highly productive machine tools. Production Engineering, Bd. 6, Nr. 3. pp 277–285
3. Dhillon B (2006) Maintainability, maintenance, and reliability for engineers. Taylor and Francis
4. DIN (2016) German standardization roadmap—industry 4.0. Version 2. Berlin
5. European Commission (2016) TOPIC: novel design and predictive maintenance technologies for increased operating life of production systems http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/fof-09-2017.html. Accessed 8 Dec 2016
6. ISO 3408-5: (2006) Ball screws—Part 5: static and dynamic axial load ratings and operational life
7. Huf A (2012) Kumulative Lastermittlung aus Antriebsdaten zur Zustandsbewertung von Werkzeugmaschinenkomponenten. Jost Jetter Verlag, Heimsheim
8. Kagermann H, Wahlster W, Helbig J (2013) Recommendations for implementing the strategic initiative INDUSTRIE 4.0
9. Lasi H, Fettke P, Kemper HG, Feld T, Hoffmann M (2014) Industry 4.0. Business and information. Syst Eng 6(4):239–242. https://doi.org/10.1007/s12599-014-0334-4
10. Lee J, Bagheri B, Kao HA (2015) A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manufact Lett 3:18–23. https://doi.org/10.1016/j.mfglet.2014.12.001
11. Lorenz M, Rüßmann M, Strack R, Lueth KL, Bolle B (2015) Man and machine in industry 4.0—how will technology transform the industrial workforce through 2025? The Boston Consulting Group
12. McKinsey&Company (2015) Industry 4.0—how to navigate digitization of the manufacturing sector
13. Ministry of Trade Industry and Fisheries (2017) The industry—greener, smarter and more innovative (in Norwegian: St.meld. nr 27 (2016–2017)—Industrien—grønnere, smartere og mer nyskapende). https://www.regjeringen.no/no/dokumenter/meld.-st.-27-20162017/id2546209/. Accessed 2 May 2017
14. Rosendahl T, Hepsø V (2013) Integrated operations in the oil and gas industry. Business Science Reference, Hershey, Pa
15. Rødseth H (2014) Maintenance optimisation for integrated planning. In: Safety, reliability and risk analysis: beyond the horizon: proceedings of the European safety and reliability conference, ESREL 2013, Amsterdam, the Netherlands, 29 Sept–2 Oct 2013. CRC Press, pp 651–657
16. Rødseth H, Schjølberg P (2016) Data-driven predictive maintenance for green manufacturing. In: Advanced manufacturing and automation VI, vol 24. Advances in economics, business and management research. Atlantis Press, pp 36–41
17. Rødseth H, Schjølberg P, Larsen LT (2016) Industrie 4.0—a new trend in predictive maintenance and maintenance management. In: EuroMaintenance 2016—conference proceedings. Artion Conferences & Events, pp 267–273

Recognition of Garlic Cloves Orientation Based on Machine Vision

Shuai Du, Yang Li, Xuemei Liu and Jin Yuan

Abstract In order to realize the garlic single grain precision seeding, keep the garlic in an upright position, ensure the state of the garlic cloves upward, this paper uses machine vision technology, the collected garlic image preprocessing, image segmentation, expansion, area recognition method to process garlic images. Using the Otsu algorithm to segment the image, we can separate the garlic image from the background very well. Based on the asymmetry of the area between the upper part and the lower part of the garlic. The direction of the garlic is identified by calculating the area of two parts. The experimental results show that the recognized accuracy of garlic's direction could reach to 95.33% by using this algorithm. And it can provide technical support for future single-grain precision planting of garlic.

Keywords Machine vision · Otsu algorithm · Garlic

1 Introduction

China is one of the world's main garlic producing and exporting countries. The area of garlic planting is more than 330 thousands hm^2. And the output of garlic is 10 million 580 thousand T, accounting for 3/4 of the world's output of garlic [1]. Moreover, garlic has rich nutritional value and medicinal value, and it has good effect on the treatment of hyperlipidemia, diabetes and other diseases [2].

At present, with the increase of demand, the planting area of garlic is increasing gradually. In recent years, a research shows that the cloves direction of the garlic planting has an important influence on the bud ratio and output of garlic, which is

S. Du · J. Yuan (✉)
College of Mechanical and Electronic Engineering,
Shandong Agricultural University, Tai'an 271018, China
e-mail: jyuan@sdau.edu.cn

Y. Li · X. Liu
Shandong Provincial Key Laboratory of Horticultural Machinery
and Equipment, Tai'an 271018, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_48

the best way by garlic planting keep upward [3]. At present, the garlic planting machinery has realized upright of the garlic planting, and ensured the upright state of garlic planting [4]. But there is the phenomenon that the end of the garlic is upward, and the garlic seeder and seeding situation are shown in Fig. 1. In recent years, some agricultural machines have joined machine vision technology [5–7]. The existing researches, such as Lei et al. [8], Yingfang and Buoyou [9], Yingfang and Zhihua [10] and Qingming et al. [11], have realized the recognition of garlic orientation by means of feature and pattern recognition. In which, Qingming et al. [11] and others [12–14] identify the garlic cloves by area, using the same height above and below the garlic cloves, and calculate the area of it. In order to realize the garlic single grain precision seeding, ensure garlic upright, the state of the garlic cloves upward, we use Otsu algorithm for image segmentation, and using the principle that the two cloves of garlic are different area, we can sever the garlic image into two areas, meanwhile by calculating the area of two parts to identify the direction of it. By applying this algorithm the vertical seeder, we can realize the garlic single grain precision seeding, ensure garlic upright, the state of the garlic cloves upward, and solving the sowing problem.

2 Image Acquisition

Garlic image acquisition as the basis of garlic clove recognition, has an important influence on the later image processing. Because of the stability of the industrial camera, in the experiment, the fixed focus CCD industrial camera was used in the experiment, the environment was natural light. After repeated experiments, it is found that the recognition rate that is photographed in the background of black rough surface (non—glossy surface) is high. Using the black matted material as the background of garlic, the pixel is set to 640 × 480 pixels.

Fig. 1 Garlic seeder and seeding condition

Under reflecting the actual garlic recognition situation, to reduce the computation complexity, and accelerate the garlic recognition process, the garlics are randomly horizontally placed with 3 postures in the black background, as shown in Fig. 2.

3 Image Preprocessing

3.1 Image Segmentation

In order to distinguish the garlic image from the background, the image threshold segmentation is needed. The threshold segmentation is to make the garlic image binarization, and give the garlic image and background to different threshold, so as to separate the garlic image from the background. An adaptive threshold algorithm is used for the threshold segmentation algorithm in the exiting research. The implementation method of the algorithm is to determine the neighborhood window with it as the center of each pixel, and calculate the maximum and minimum of the pixels in the window, and then take their mean value as the threshold. Considering the actual situation, under the influence of natural light intensity, the garlic image taken by industrial camera will also be different, always compared by uncertainty, however, using Otsu algorithm for image segmentation has a high level of stability, and avoid this problem very well. The accuracy of image segmentation will directly affect the later recognition of garlic flap, that is, the image can be segmented completely, which will greatly enhance the accuracy of garlic clove recognition.

In this paper, Otsu algorithm is used for image segmentation of garlic. The threshold of Otsu method is based on the idea that the gap between the target and background should be the biggest.

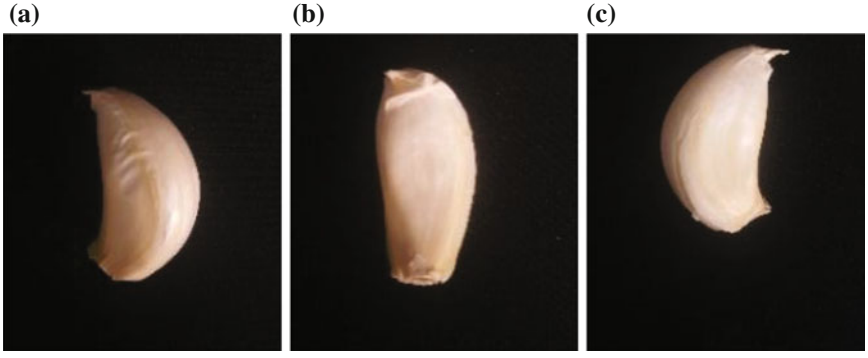


Fig. 2 Three poses of garlic

Threshold formula:

$$T(x) = \alpha_0 \cdot \alpha_1 \frac{[P_0(x) - P_1(x)]^2}{\partial_0(x) + \partial_1(x)} \tag{1}$$

where

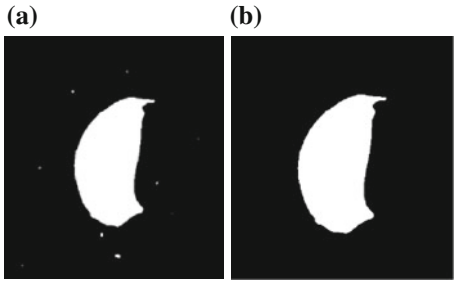
$$\partial_0(x) = \frac{\sum_{i=0}^{x} [i - P_0(x)]^2 P(i)}{\omega_0}, \quad \partial_1(x) = \frac{\sum_{i=x+1}^{L-1} [i - P_1(x)]^2 P(i)}{\omega_1} \tag{2}$$

and where, $\partial_0(x), \partial_0(x)$ are variance, $P_0(x)$ is target mean value, $P_1(x)$ is background mean value, α_0 is proportion of target parts, α_1 is proportion of background parts, ω_0, ω_1 are introducing variance.

Compared with other image segmentation algorithms, Otsu algorithm has strong adaptability. With the light intensity of the environment, the transformation of garlic size can adjust the threshold itself, has better anti-jamming ability, image segmentation is more stable. The garlic image segmented by Otsu algorithm is shown in Fig. 3a.

3.2 Image Expansion

After segmentation of garlic image by Otsu algorithm, a small amount of white spots appeared around the garlic image. The appearance of the white spot affects the segmentation and processing of garlic image, it needs to be expanded, by filling the surrounding white spots into garlic image. Using the structural element of 3 × 3, each pixel of the image is scanned, the structure image is used to do "or" operation with the covered binary image, the result is 0, and the pixel of the image is 0, otherwise it is 1. The expanded image is shown in Fig. 3b.



Segmentation of garlic images | Garlic images after expansion

Fig. 3 Garlic processing image

4 Garlic Flap Recognition

4.1 Garlic Characteristic Parameters

The existing recognition rate of garlic clove recognition algorithm which has high recognition rate using the SUSAN corner detection algorithm. The recognition rate can reached 95%. The idea of the algorithm is that through the class circle template moving on the image, the gray value of each image pixel in the template is compared with the gray value of the template center pixel, if the difference between the gray level of a pixel in the template and the center pixel of the template is less than a certain threshold, it can be considered that the point and the nucleus have the same gray level, thus the region satisfying the formation of the condition is called the nuclear value similarity region (USAN). When the template moves to the tip of the garlic cloves, the value of USAN is the smallest, the point is the tip of garlic.

After the observation of garlic, each cloves garlic is irregular, but one thing in common is that the top area of garlic clove is obviously smaller than the end of garlic tail, as shown in Fig. 2. In this paper, the area of the tip and tail part of the flap is used as the recognition of garlic petals [12].

4.2 Recognition of Flap Direction Based on Area Feature

The garlic image, after the previous processing, has completely been separated from the background. The idea of the area calculation algorithm of disc tip and flap tail is that:

(1) The edge of the segmented garlic image is traced, and the boundary of garlic and the coordinate value of each pixel point are obtained. $X_{pixl}[i], Y_{pixl}[i]$ (i-represents the number of pixels).
(2) Considering the characteristics of each pixel in the image, using freeman chain code to express the pixel boundary in 8 directions, the Freeman chain code representation is shown in Fig. 4.

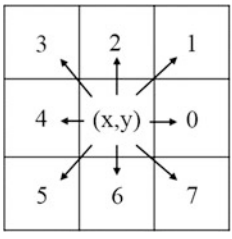


Fig. 4 Freeman chain code representation

(3) Make 2 vector labeling of the boundary pixel of garlic, respectively, pre-vector $S_{\text{front}}[i]$ and post-vector $S_{\text{behind}}[i]$ labeling:

$$S_{\text{front}}[i] = \begin{cases} -1 & i > 0, \quad Y_{pixl}[i] > Y_{pixl}[i-1]; \quad i = 0, \quad Y_{pixl}[0] > Y_{pixl}[N-1] \\ 0 & i > 0, \quad Y_{pixl}[i] = Y_{pixl}[i-1]; \quad i = 0, \quad Y_{pixl}[0] = Y_{pixl}[N-1] \\ 1 & i > 0, \quad Y_{pixl}[i] < Y_{pixl}[i-1]; \quad i = 0, \quad Y_{pixl}[0] < Y_{pixl}[N-1] \end{cases} \tag{3}$$

$$S_{\text{behind}}[i] = \begin{cases} -1 & i < N-1, \quad Y_{pixl}[i+1] > Y_{pixl}[i]; \quad i = N-1, \quad Y_{pixl}[N-1] > Y_{pixl}[0] \\ 0 & i < N-1, \quad Y_{pixl}[i+1] = Y_{pixl}[i]; \quad i = N-1, \quad Y_{pixl}[N-1] = Y_{pixl}[0] \\ 1 & i < N-1, \quad Y_{pixl}[i+1] < Y_{pixl}[i]; \quad i = N-1, \quad Y_{pixl}[N-1] < Y_{pixl}[0] \end{cases} \tag{4}$$

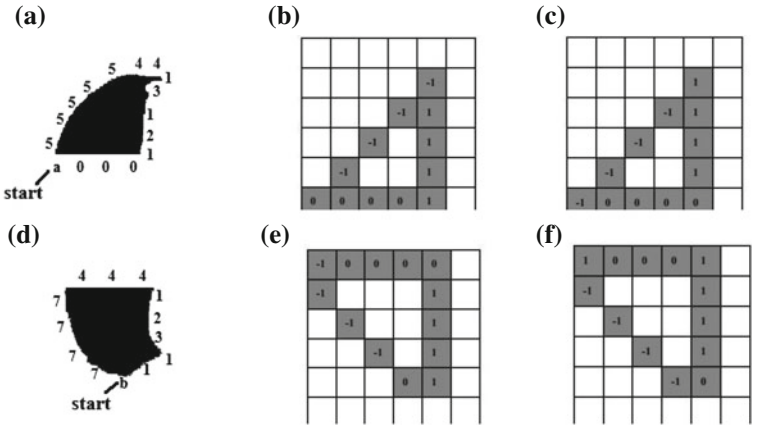
The border tracking of the upper part of garlic, starting at point "a", as shown in Fig. 5a, the next half of the garlic border tracing, starting point "b", as shown in Fig. 5d. The garlic image is labeled with a vector, as shown in Fig. 5.

The completed vector is regularly processed to obtain the final vector $S[i]$. Finally, the area of garlic image is calculated

$$S = \sum_{i=0}^{N-1} X_{pixl}[i] \cdot S[i] + N_1 \tag{5}$$

where N_1 is the number of $S[i]$ equals 1.

The calculation process is: by line scanning, the iterator accesses the image pixel points, calculates the pixel points of the gray value which is 255, calculates the



a&d. Image boundary of garlic **b&e.** Pre-vector annotation **c&f.** Post vector annotation

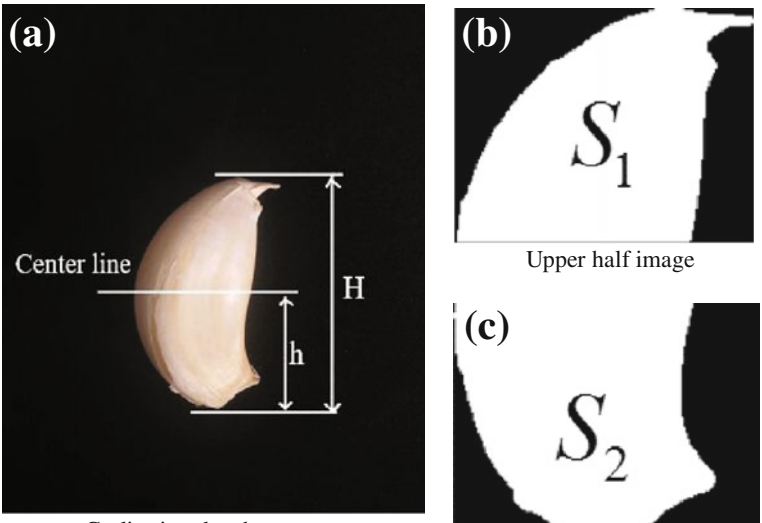
Fig. 5 Image and vector annotation of garlic

height h of garlic in the image, and also calculates the height h of the garlic image center line through the height h, as shown in Fig. 6a. The garlic image is divided into the upper and lower parts by the center line, as shown in Fig. 6b and c, then the total area of the upper and lower parts are calculated respectively, and the two parts are compared later. By calculation, we can see that $S_1 < S_2$. The upper part is the garlic clove tip, and return to the garlic petal tip upward.

5 The Results and Analysis of Experiment

Nowadays, widely planted garlic varieties are Cangshan garlic, Jinxiang garlic, Qi County garlic, those three kinds of garlic have a certain difference in appearance and particle plumpness. According to the agricultural planting requirements of garlic, the garlic used as garlic must be pure skin, shape specification, full, and skin completed. In order to verify the reliability and versatility of the flap recognition algorithm, the diameter of garlic in the above three varieties is selected in each plump 100 grains, and the three kinds of poses shown in Fig. 2 are arranged randomly. The collected garlic image is processed by the flap recognition algorithm, and the identification results are shown in Table 1.

According to the recognition results of garlic cloves in the table, it can be seen that the overall recognition rate of garlic cloves by this algorithm is high, among which the three varieties of garlic, especially Cangshan garlic and Jinxiang garlic



Garlic size sketch map

Upper half image

Lower half image

Fig. 6 Garlic size and tip and tail recognition image

Table 1 Recognition results of garlic cloves

Garlic varieties	Cangshan garlic	Jinxiang garlic	Qixian garlic
Total image	100	100	100
Correct recognition	96	95	95
Recognition rate %	96	95	95
Time consumption/s	0.2411	0.2150	0.2560

have high recognition rate. The recognition rate of garlic is closely related to the plumpness of garlic granules and the color of it. Using the algorithm to identify the part direction, the overall recognition rate can reach 95.33%, and the average time consumption is 0.2374 s.

6 Conclusion

(1) This paper uses machine vision technology, according to the asymmetry of the two parts of garlic tip and tail area, choosing area as the feature of the petal to identify. The experimental results show that the image processing method based on machine vision is definitely feasible in identifying the direction of garlic cloves.
(2) In this paper, Otsu algorithm is used for image segmentation, the garlic can be well separated from the background, adopting the method of dividing garlic image into two parts, and directly through Freeman chain code method to calculate the upper and lower parts of the area, to improve the accuracy of garlic clove recognition.
(3) In the experiment, the three kinds of garlic cloves recognition was compared, and the universality and reliability of the garlic flap to the algorithm were increased, providing technical support for the garlic single grain precision seeding.

References

1. Xiaoyong X, Chongguang L (2013) Study of "Power Effect" in China's Garlic Exports. J Int Trade 08:61–71. https://doi.org/10.13510/j.cnki.jit.2013.08.008
2. Yongcheng Y (2016) Nutritive value of garlic and high yield and high efficiency technology. Agri Technol (12):36. doi 10.11974/nyyjs.20160632032
3. Chengqian J, Wensheng Y, Chongyou W et al (2008) Experimental study on effects of the bulbil direction on garlic growth. Trans CSAE 24(4):155–158
4. Yi G, Zuli Z, Liying Y, Xudong Z (2009) The Present situation of the garlic planting machine. J Agri Mech Res (06):221–223. doi 10.13427/j.cnki.njyi.2009.06.028

5. Camargo A, Smith JS (2009) An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosyst Eng 102(1): 9–21. ISSN 1537-5110, https://doi.org/10.1016/j.biosystemseng.2008.09.030
6. Zhang B, Huang W, Gong L, Li J, Zhao C, Liu C, Huang D (2015) Computer vision detection of defective apples using automatic lightness correction and weighted RVM classifier. J Food Eng 146:143–151. ISSN 0260-8774. https://doi.org/10.1016/j.jfoodeng.2014.08.024
7. Blok PM, Barth R, van den Berg W (2016) Machine vision for a selective broccoli harvesting robot. IFAC-PapersOnLine 49(16):66–71. ISSN 2405-8963, http://dx.doi.org/10.1016/j.ifacol.2016.10.013
8. Lei P, Shijian D, Ronghua L (2010) Research on the identification of the roots of garlic base on pattern recognition. J Agri Mech Res 05:51–54. https://doi.org/10.13427/j.cnki.njyi.2010.05.016
9. Yingfang G, Buoyou L (2011) Research on the identification of the roots of garlic based on SUSAN algorithm. J Yang Ling Vocat Tech Coll (03):1–4
10. Yingfang G, Zhihua W (2015) Recognition of garlic clove tip based on image processing. Shanxi J Agri Sci 08:54–57
11. Qingming Y, Juanling L, Ruiyin H (2010) Direction identification of garlic seeds based on image processing. Acta Agriculturae Zhejiangensi 01:119–123
12. Wang Y (2012) Research and application of image recognition method base on visual feature distribution. Dalian University of Technology
13. Chi G, Hui G (2013) Direction identification system of garlic clove based on machine vision. TELKOMNIKA 11(5):2323–2329
14. Yang S, Li T, Xiao S, Zhou H, Xu N (2015) Research on improved Otsu algorithm and its application in moving objection detection. Video Appl Project (24):100–103 + 112

Metal Penetration in Additively Manufactured Venting Slots for Low-Pressure Die Casting

Even Wilberg Hovig, Knut Sørby and Per Erik Drønen

Abstract Venting of excess gas and air in casting moulds is of critical importance to avoid defects such as misruns and porosities. Additive manufacturing give mould designers a possibility to make venting slots in large areas and geometrically complex mould sections. The venting slots must be wide enough to allow air to escape, but narrow enough to restrict metal flowing into the venting slots. In this paper, an experiment was conducted to identify which process parameters cause metal penetration in venting slots in low-pressure die casting (LPDC). Parameters having the largest influence on the degree of penetration is found to be filling pressure and slot width. For a typical LPDC process a slot width of 0.14 mm is recommended.

Keywords Low-pressure die casting · Additive manufacturing
Venting

1 Introduction

The purpose of the research presented in this paper is to investigate which parameters cause molten aluminium A356 (Al7Si0.4Mg) to penetrate additively manufactured venting slots in low-pressure die casting (LPDC) tools.

E. W. Hovig (✉) · K. Sørby
Department of Mechanical and Industrial Engineering, NTNU Norwegian University of Science and Technology, Trondheim, Norway
e-mail: even.w.hovig@ntnu.no

P. E. Drønen
Benteler Automotive Farsund AS, Farsund, Norway

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_49

1.1 Additive Manufacturing

Additive manufacturing (AM) has grown from mainly being used as a tool for rapid prototyping to the manufacture of end-use products, including tooling components. As of 2016 7.2% of all AM components are used in tooling [1], up from 4.3% in 2013 [2]. AM is a new tool for the die-maker and can potentially increase process-control in a casting process. The benefits include an increased flexibility for the tool-maker in addition to increased thermal control [3], and the manufacturing of moulds is faster [4]. Another potential benefit is the ability to evacuate the excess gas and air in the mould cavity through venting slots in AM inserts in a permanent metal mould.

AM has seen extensive use in injection moulding [1] which benefits from the possibilities given by conformal cooling channels and complex internal geometries. Work has been done regarding the use of AM materials for aluminium high-pressure die casting (HPDC), and the results have been promising [5]. The mould material properties requirements are comparable in HPDC and LPDC, and the wear of the mould in a HPDC process will generally exceed the wear of a mould in a LPDC process with comparable process parameters.

1.2 Low-Pressure Die Casting

LPDC is used to produce parts with high integrity and in large volume. Typical examples are aluminium castings for the automotive industry. Sand cores or metal cores can be incorporated in the process to produce hollow parts.

In low-pressure die casting the molten metal is forced into the mould through a fill tube that is immersed in a crucible below the mould. The air around the crucible is pressurized to generate the necessary force to lift the molten metal into the mould. A schematic illustration of a low-pressure die casting machine is shown in Fig. 1. A distinct advantage of the process is that clean molten metal is introduced into the fill tube, and surface oxides from the melt are avoided in the mould.

As the metal fills the mould, the air in the mould, and any gas from the sand core binder must be evacuated through properly designed venting elements. If excess gas are not properly evacuated in a LPDC process defects such as misruns and porosities may occur. The defects could be found on the surface of the casting, or inside the casting. A key feature of the venting slot is to allow air to flow through, but not molten aluminium.

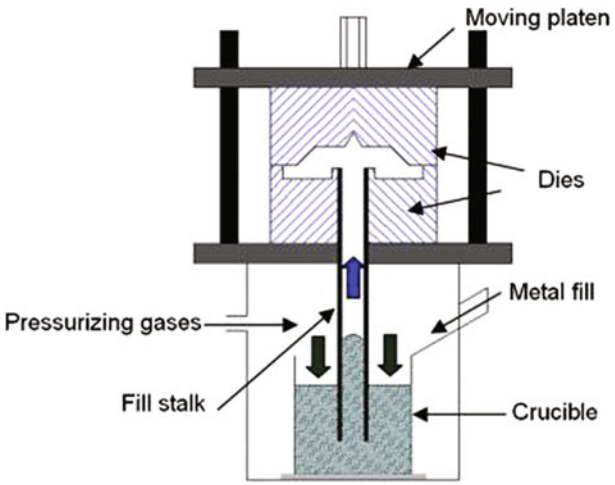


Fig. 1 Low-pressure die casting machine [6]

1.3 Test Parameters

Key parameters are identified as vacuum level, casting temperature (molten metal temperature), mould temperature, contact time, and slot size. Vacuum level and casting temperature are tested at two levels. The target mould temperature is 300 °C, and the target contact time is between 5 and 10 s.

1.4 Surface Tension and Oxide Films

Aluminium does not typically wet moulds, which results in a capillary back-pressure restricting the flow of the melt [7]. The back-pressure, P_{ST}, is modeled after the Young-Laplace equation as shown in (1), where γ is the surface tension, and r and R are the principal radii of curvature of the meniscus of the melt.

$$P_{ST} = \gamma(1/r + 1/R) \tag{1}$$

In a case where the melt is filling a thin wide slot, where $R \gg r$, (1) can be rewritten as $P_{ST} = \gamma/r$. In this case r can be expressed as $b/(2*\cos(\alpha))$, where α is the contact angle and b is the slot width. Assuming zero wettability between the aluminium and mould the contact angle is 180°, and r equals half the slot width. The back-pressure is then a function of the slot width and the surface tension. The effective surface tension of molten A356 is highly dependent on the surface oxide layer and temperature. Reported values of surface tension for A356 is in the range of 0.390–0.845 N/m [8, 9].

The aluminium A356 alloy used in this experiment is modified with Strontium, which is known to reduce the surface tension.

2 Experiment

2.1 Degree of Metal Penetration

Two sets of five disks were built in two different powder bed fusion (PBF) machines, as shown in Table 1, to simulate a casting mould. Each disk had six groups of venting slots with 8 slots in each group. The nominal slot widths of each venting slot group is given in Table 1. The diameter of the test disk is 80 mm and the thickness is 5 mm. The disks were built with the slots aligned with the build direction, as to minimize the overhang in the AM process.

Each disk was coated with a graphite coating and placed in a testing apparatus as shown in Fig. 2. The disk was first preheated with a propane burner before it was lowered a set distance, d, into a crucible of molten aluminium A356 at temperature T for a duration of t seconds. A vacuum was applied to the chamber above the disk with a pressure of P mbar as the disk was lowered into the molten aluminium. The oxide layer on the top surface of the crucible was removed immediately before the disk was lowered into the melt pool. The test parameters are summarized in Table 2. The casting temperature and vacuum was controlled by a LPDC machine, and the disk temperature was measured by a thermocouple placed in the center of the disk.

The degree of metal penetration was determined as the average percentage of the slot area in each slot-group filled with aluminium. The area was determined by capturing an image of each disk and counting how many pixels the aluminium covered in the longitudinal direction of each slot. This value was divided by the total number of pixels in the longitudinal direction of the slot.

Table 1 Disk properties

Disk	AM process	Material	Nominal slot widths (mm)
1 through 5	Direct metal laser sintering (EOS EOSINT M280)	Steel 1.2709 (Vendor name MS1)	0.10 0.12 0.14 0.16 0.18 0.20
6 through 10	Laser cusing (Concept laser M2)	Steel 1.2709 (Vendor name CL50WS)	0.16 0.18 0.20 0.22 0.24 0.26

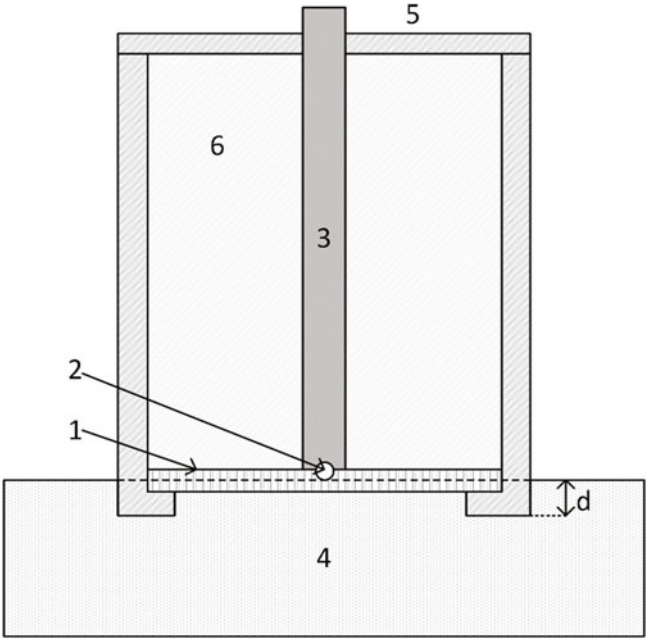


Fig. 2 Experimental set-up. The test disk (1) is fixed in a cylindrical holder and lowered a distance (*d*) into the melt pool (3). The temperature of the disk is measured with a thermocouple (1) whose wires runs through a steel tube insulated with a ceramic material (2). The chamber above the disk (5) is depressurized with a vacuum pump connected to the top plate (4)

Table 2 Test parameters

Disk	$T_{casting}$ (°C)	T_{disk} (°C)	P_{target} (mbar)	P_{peak} (mbar)	Contact time, t (s)
1	650	297	50	50	7
2	650	272	100	160	8
3	700	300	50	50	7
4	700	295	100	121	9
5	700	236	50	69	10
6	650	303	50	52	6
7	650	299	100	156	8
8	700	303	100	118	9
9	700	296	50	62	7
10	700	287	25	50	10

2.2 Actual Slot Size

The actual slot size was measured on one disk from each PBF machine in a microscope. Each slot was measured three times, at the bottom, middle, and top of each slot, giving a total of 24 measurements for each slot group.

The actual slot widths differs substantially from the nominal slot widths. As seen in Table 3, the deviations of actual slot widths compared with nominal slot widths are in the range of 0.003–0.069 mm. Perhaps more interesting than the deviation, from an AM point of view, is how the slot widths are grouped. Considering disk 1, the slots with nominal widths 0.12, 0.14, and 0.16 mm all have actual slot widths of about 0.14 mm. The same is found in disk 6 for nominal slot widths 0.16, 0.18, and 0.20 mm. In this case it is grouped around 0.18 mm.

In a PBF process a metal powder is melted with a high power laser beam and the characteristics of the material changes dramatically. On a powder scale, particles on the vicinity of the melt pool will only be partially melted. This is problematic at the contours of holes and slots, where no overlap of the laser beam is possible. For slots with nominal widths of 100–260 µm partially melted powder particles with diameters typically in the range of 10–70 µm can have a dramatic effect on the actual slot width.

3 Results and Discussion

3.1 Metal Penetration

The degree of metal penetration is shown in Fig. 3. The disk number (from Table 1) is indicated on the vertical axis. The slot size is indicated on the horizontal axis.

Table 3 Actual slot widths in mm

	Disk 1					
Nominal slot width	0.10	0.12	0.14	0.16	0.18	0.20
Average actual slot width	0.097	0.147	0.135	0.138	0.160	0.166
Deviation from nominal width	−0.003	0.023	−0.005	−0.022	−0.020	−0.034
Standard deviation (24 measurements per slot group)	0.014	0.057	0.022	0.012	0.025	0.030
	Disk 6					
Nominal slot with	0.16	0.18	0.20	0.22	0.24	0.26
Average actual slot width	0.180	0.183	0.187	0.244	0.236	0.329
Deviation from nominal width	0.020	0.003	−0.014	0.024	0.004	−0.069
Standard deviation (24 measurements per slot group)	0.020	0.040	0.043	0.034	0.037	0.066

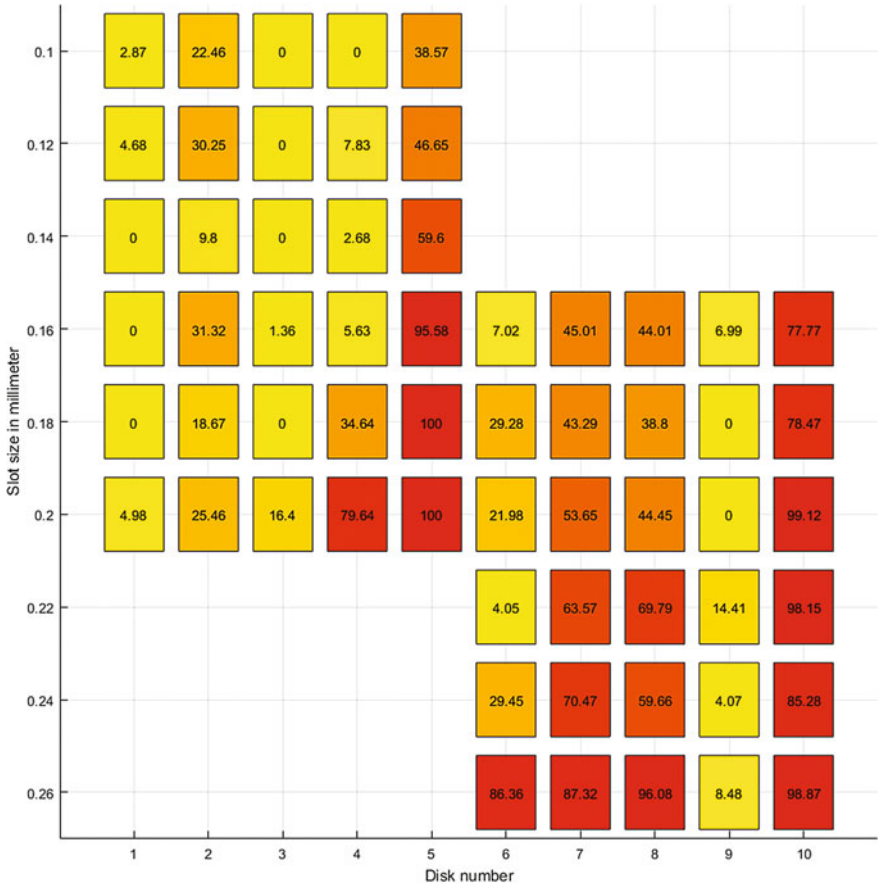


Fig. 3 Degree of metal penetration as a function of slot size. The degree of metal penetration as a percentage is indicated both as a numerical value and by the color legend

Contact Time

Considering disk 1, 3, 5, 6, and 9, which had a target vacuum level of 50 mbar there was a low degree of metal penetration for all slot sizes, with the exception of disk 6–0.26 mm slot width and all of disk 5. Disk 5 was held longer in the melt pool (10 s) compared to the other disks considered (6–7 s). The temperature of disk 5 was lower (236 °C) compared with disk 1, 3, 6, and 9 ($\sim$300 °C). Disk 10 also had a contact time of 10 s, and a higher degree of penetration compared with the other disks with a low target vacuum. As the aluminium comes in contact with the cooler steel disk a thin layer of aluminium is immediately cooled down and solidified. As the contact time increases the solidified aluminium re-melts, and penetration is again possible. In a casting process, however, the melt pool has a significantly lower mass compared with the mould, and re-melting of the aluminium is not considered an issue.

Actual Slot Size

There is a significantly higher level of metal penetration in the comparable slots in disk 6 compared to disk 1. However, the actual slot sizes differ substantially from the nominal slot sizes, and thus explains the elevated level of penetration.

Casting Temperature

To look at the effect of casting temperature disk 1 and 3, disk 2 and 4, disk 6 and 9, and disk 7 and 8 is compared pairwise. For these experiments the vacuum level for each disk pair is comparable, but the casting temperature is either 650 or 700 °C.

For the low vacuum case (disk 1 and 3, and disk 6 and 9) there is little difference in melt penetration for disk 1 and 3. Considering disk 6 and 9 there is a significantly higher degree of penetration in disk 6, although the process parameters for disk 9 should indicate a higher degree of penetration in disk 9. The low level of penetration in disk 9 is most likely a result of a thicker surface oxide film layer on the melt pool effectively creating a physical barrier stopping the flow into the venting slots.

For the high vacuum case with similar disk temperature (disk 7 and 8), there is no significant difference in melt penetration. In the other high vacuum case (disk 2 and 4) there is a higher degree of penetration in disk 2 compared to disk 4. In disk 4 the disk temperature is slightly higher (23 °C) and the contact time is higher. This will increase the chance of re-melting, as discussed previously. The low level of penetration in disk 4 can also be a result of a thicker surface oxide layer due to a failure to remove it from the melt pool before the disk was lowered into the aluminium.

The casting temperature will affect the surface tension of the aluminium, but a temperature difference of 50 °C is considered to have negligible impact on the Young-Laplace Eq. (1) when compared to the larger pressure difference of 50 mbar.

Vacuum Level

Disk 2, 4, 7, and 8 had a target vacuum of 100 mbar and peak values in the range of 118–160 mbar. The degree of metal penetration was generally higher for the experiments with the higher vacuum level. This further suggests that vacuum has a higher impact on melt penetration compared to casting temperature.

Mould Temperature

Disk 5, with a low initial temperature had a high degree of metal penetration, even though the vacuum level was low compared to disk 4. This suggests that the mould temperature is of significance. This can also be seen in the pattern of solidification where penetration occurred. In disk 2, 5, 7, and 9 there is a distinct circular pattern of penetration, exemplified in Fig. 4.

The steel tube holding the thermocouple acts as a chiller and any melt penetrating the slot will solidify faster in the center region, thus not being able to retract as discussed in the next section.

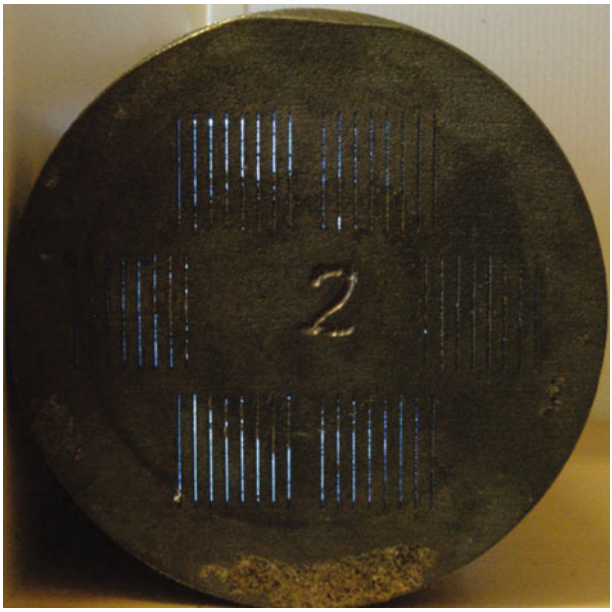


Fig. 4 Circular pattern of penetration in disk 2

3.2 Metal Retraction

Figure 5 shows the solidified aluminium after disk 1 was tested. It is clear that the melt enters the slot, but is released after the casting is removed from the mould. As long as the melt does not pass through the entire slot this poses no problem in respect to clogging, but the surface quality of the cast part is compromised. It can

Fig. 5 Close up of the solidified aluminium at the bottom of disk 1 after the experiment was conducted. The melt enters the venting slot, but is released from the slot after the solidified casting is removed from the mould

also restrict the ejection of the cast part from the mould, but this problem can be solved by implementing a draft angle in the design of the slot. Figure 6 shows a piece of aluminium separated from the top of disk 5, where the melt was allowed to pass through the slot. This will cause clogging of the slots and effectively negate the effect of the venting slot.

It is interesting in Fig. 6 to notice the pattern of penetration through the slot. It is not a conform penetration in a rectangular pattern, but rather a series of points where the metal flows through each slot. This can be attributed to the surface roughness of the slots, but also because of the oxide layer on the surface of the melt. Although measures were taken to reduce the surface oxide layer thickness, the effects are considered impossible to eliminate. In Al7Si0.4 Mg alloys a thin surface film is formed almost immediately. This surface film can be as thin as 20 nm [10], but still effect the flow. The initial film to form on aluminium alloys with more than 0.005%Mg is $MgAl_2O_4$, which has a higher melting point than the aluminium alloy. The high melting point (2135 °C) will cause the oxide film to act as a physical barrier the aluminium melt has to penetrate before it can enter the slot. For such small areas the oxide film seems to temporarily restrict the flow before it splits open allowing the melt to be jetted through.

3.3 Surface Tension and Metal Penetration

Comparing the experimental results with the Young-Laplace Eq. (1), in the lower bound of surface tension (0.390 N/m) penetration occurs for slot widths larger than 0.14 mm for the low vacuum level, and all slots with the high vacuum level. In the higher bound of surface tension penetration will only occur for the high vacuum level and slot widths larger than 0.16 mm.

Fig. 6 A piece of aluminium which penetrated the 0.26 mm slot width group on disk 5 and solidified in the vacuum chamber

Considering disk 1, neglecting the low levels of penetration, the surface tension is estimated to be less than 0.5 N/m using (1). Considering disk 4 where considerable penetration appeared in slot widths above 0.18 mm the peak vacuum level was 121 mbar. According to (1) the surface tension is in the upper bound with a value of 0.9 N/m. A detailed study of the surface tension is out of the scope of this paper, but it is worth noting that there is significant variance in the observed surface tension in this experiment. This in turn is related to the varying thickness of the surface oxide layer and the surface roughness of the slots, and does not necessarily reflect the true surface tension.

4 Summary and Conclusions

The main factors influencing the degree of penetration of molten A356 in venting slots in permanent moulds in low-pressure die casting are pressure difference and slot width.

Casting temperature and surface oxide layer thickness effects penetration through altering the surface tension. A higher surface tension reduces the risk of penetration. This is seen both in the experimental results and from the Young-Laplace model of capillary back-pressure.

A high contact time between molten aluminium and mould can cause re-melting of initially solidified aluminium and penetration may occur. The mould designer must consider this in high temperature areas of the mould, e.g. in close proximity to the fill stalk.

The research presented in this paper shows that AM can be successfully used to manufacture moulds for low-pressure die casting with venting slots for excess gas evacuation. The authors recommend venting slots with 0.14 mm slot widths for a typical LPDC process.

Acknowledgements The authors thank the Research Council of Norway for supporting this work through the NextForm project, Grant No. 235241.

The authors thank Benteler Automotive Farsund AS for sponsoring the research presented in this paper in the form of providing access to LPDC machines and facilitating the execution of the experiments.

The authors thank Tronrud Engineering for supplying test disks 1 through 5, and Olav Åsebø Berg at SINTEF Raufoss Manufacturing AS for designing the test disks and supplying disks 6 through 10.

References

1. Wohlers T, Caffrey T, Campbell I (2016) Wohlers Report 2016: 3D printing and additive manufacturing state of the industry annual worldwide progress report. Wohlers Associates, Inc

2. Wohlers T, Caffrey T (2013) Wohlers Report 2013: Additive manufacturing and 3D printing state of the industry annual worldwide progress report. Wohlers Associates, Inc
3. Hovig EW, Brøtan V, Sørby K (2016) Additive manufacturing for enhanced cooling in moulds for casting
4. Brøtan V, Berg OÅ, Sørby K (2016) Additive manufacturing for enhanced performance of molds. Procedia CIRP 54:186–190
5. Pereira M, Williams M, Du Preez W (2012) Application of laser additive manufacturing to produce dies for aluminium high pressure die-casting. S Afr J Ind Eng 23(2):147–158
6. Fu P et al (2008) Low-pressure die casting of magnesium alloy AM50: response to process parameters. J Mater Process Tech 205(1):224–234
7. Ravi KR et al (2008) Fluidity of aluminum alloys and composites: A review. J Alloy Compd 456(1):201–210
8. Bainbridge IF, Taylor JA (2013) The surface tension of pure aluminum and aluminum alloys. Metall Mater Trans A 44(8):3901–3909
9. Anson JP, Drew RAL, Gruzleski JE (1999) The surface tension of molten aluminum and Al-Si-Mg alloy under vacuum and hydrogen atmospheres. Metall Mater Trans B 30 (6):1027–1032
10. Campbell J (2015) Complete casting handbook: metal casting processes, metallurgy, techniques and design, 2nd ed. London, GBR: Elsevier Science, London

Analysis and Impact Assessment in Sustainable Industrial and Infrastructural Development Projects

Lea Kristine Myklebust and Odd Myklebust

Abstract It is generally understood that large human development activities have impact on the natural ecosystems. Such impact can in many cases lead to negative environmental and climate consequences. This paper gives an overview of ecosystem impact combined with circular economy in an industrial perspective and the importance of keeping the diversity in nature in future industrial development. It includes a Norwegian industrial example with proof of practice of how small adjustments can reduce negative impact on environment in industrial and infrastructural development projects. Consequence analysis will shed light on various impacts. When unfortunate or risky consequences have surfaced through systematic analyses, it may be easy to pay immediate attention to ecosystems. Minor measures can reduce both short term and long-term negative effects on the ecosystem. We can observe that the importance of environmental issues today is very seriously treated both by United Nations (UN) and by the world countries in the Paris Agreement.

Keywords Sustainable industry · Ecosystems · Circular economy
Product Life-cycle management

1 Introduction

The United Nations Development Program (UNDP) has been one of the leading organizations working to achieve the Millennium Development Goals (MDGs). UNDP is present in more than 170 countries and territories. There are all together 17 goals in the program. In the context of this paper, five of them are relevant for the issues in this paper [6].

L. K. Myklebust (✉)
Norwegian University of Life Sciences (NMBU), 1433, Ås, Norway
e-mail: lkmyklebust@gmail.com

O. Myklebust
SINTEF Raufoss Manufacturing, 7031 Trondheim, Norway
e-mail: Odd.Myklebust@sintef.no

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_50

Goal 9 Industry Innovation and Infrastructure focuses on the promotion of infrastructure development, industrialization and innovation (Fig. 1).
Goal 12 Responsible Consumption and Production aims at specific policies and international agreements on the management of materials that are toxic to the environment.
Goal 13 Climate actions; Climate change presents the single biggest threat to development, and its widespread, unprecedented effects disproportionately burden the poorest and the most vulnerable peoples.
Goal 14 Life below Water seeks to promote the conservation of and sustainable use of marine and coastal ecosystems, prevent marine pollution and increase the economic benefits to small islands developing States and LDCs from the sustainable use of marine resources.
Goal 15 Life on Land focuses on managing forests' sustainably, restoring degraded lands and successfully combating desertification, reducing the ongoing degradation of natural habitats and ending biodiversity loss [6].

Climate change is accelerating and the Paris agreement shows that most of the countries in the world take climate changes seriously. Norway has signed the Paris agreement, and is thereby obliged to reduce own CO_2 emissions by 40% within 2030 based on the Emission levels 1990 and be a low emission society within 2050 —Despite the obligation to reduce emissions, only a few politicians suggest a budget that should lead to actual reduction of emissions. The People's Republic of China has also signed the Paris agreement along with 175 other countries [14]. China has thus committed to the following actions by 2030 [1]:

- Peaking of carbon dioxide emissions around 2030 and making best efforts to peak early
- Lowering carbon dioxide intensity (carbon dioxide emissions per unit of GDP) by 60–65% from the 2005 level

Fig. 1 United Nations, sustainable development with 17 goals [6]

- Increasing the share of non-fossil fuels in primary energy consumption to around 20%
- Increasing the forest stock volume by around 4.5 billion cubic meters from the 2005 level.

2 Industry in Biodiversity and Ecosystems

Infrastructure (roads, railroads, airports, harbours etc.) or industrial facilities (process plants, factories, power plants etc.) constructions are all operating within the nature ecosystem (wetland, forests, lakes, mountains etc.). This must be done in the sustainable way. Although extinction is a natural phenomenon, it occurs at a natural "background" rate of about one to five species per year. Scientists estimate we're now losing species at 1000 to 10,000 times the background rate, with literally dozens going extinct every day [5]. However, any human industrial project has to handle the double perspective;

- How can the environment or nature threaten or damage the human project (weather condition, inundation flood, noxious animals etc.)?
- How can the human project threaten or damage the nature habitats and environment (air emissions, toxic waste, radioactivity etc.)?

At least 40% of the world's economy is derived from biological recourses [9]. Industrial facilities (power plants, process plants, factories, oil & gas installations etc.) and infrastructure (roads, railroads, airports etc.) are all examples of constructions that influence, degrade, and often annihilates natural habitats like forests, rivers and creeks with all their living organisms. Forest reserves, national parks, building restrictions etc. are often seen as extra costs for society compared to using the area for timber production or building a new factory. However, this is not necessarily the case, as natural habitats provide the ecosystem services. The values of these services do not benefit just shareholders and the landowners, but to all stakeholders in society.

Moors and wetlands benefit human society enormously as they provide safe and secure handling of floodwater, cleanse water, and store carbon. In fact, it is estimated that, although covering only 5–8% of the terrestrial landscape, wetlands store an estimate of 300–700 billion tons of carbon globally. The existing storage of carbon in the wetlands approaches the amount of carbon there is in the atmosphere [7]. Saving marshland can balance budgets in a social economic perspective. Another aspect is the importance of bees, which is well known. We have to deal with their need for flowers, so that pollination will continue to happen the natural way. All ecosystems have their respective properties. Knowledge of mapping of plants, flowers, insects, vertebrates and other organism groups, especially the red listed ones is essential. Considering ecosystems at an early stage in the industrial projects may promote sustainable solutions. Important information from local

people and organizations must be taken into account. Such knowledge, attitudes and systematic work are important for our future.

Examples of solutions, which can reduce the impact on natural habitats, are ecosystem-based evaluation of alternative locations and solutions for industrial and infrastructure facilities, e.g. construction of bridges or tunnels for safe road animal crossings, reduced emissions etc. Industrial solutions that involve environmental issues are often considered more costly. Crossing tunnels for small animals, such as frogs, might sound banal, and is an extra cost to a road project, but can be crucial for the survival of an animal population. However, such environmental strategies are often referred to as compensating measures of ecological compensation.

The industrial facilities are in all cases making a product or a service (included energy). The industry facility also has a lifecycle and in that respect all the lifecycle phases are a subject for impact to nature, climate and environment (Fig. 2).

The important impact factors for sustainable industries can be as follows:

- Choice of location and size
- Methods of facility construction
- Material consumption e.g. use of hazard materials
- Waste, emissions in operation
- Refurbishment and decomposition.

Handling waste does not necessarily generate cost. Waste for one is a resource for someone else and that is one key point in circular economy [12].

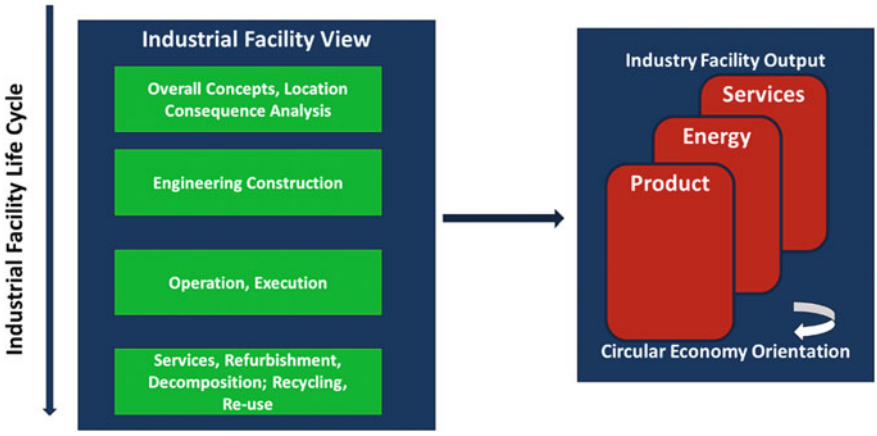


Fig. 2 Global sustainable industrial facility view

3 Consequence and Risk Analyses

There are many models for consequence analysis. A life-cycle assessment and a risk assessment for industrial facilities are needed to cover the industrial as well as ecological challenges [10].

Consequence analyses focus on;

- Options for location, clean production, energy aspects
- Priority of nature issues and values
- Consequences for natural habitats, like habitats for birds, amphibians, mammals etc.

In addition, the three main aspects of sustainability and life cycle analyses and impact have to be included;

- Environmental
- Economic
- Social.

Industrial and infrastructural initiatives or projects intervene in nature and landscape, local environment, outdoor life etc. This can, in the next phase imply reduction in biodiversity, recreational areas, harvesting areas, changes in the conditions of existing culture environment and natural resources. There are both priced and not priced consequences and values.

- In a priced consequence philosophy (quantitative issues) the costs for the society are based on the willingness to pay for the consequences of the initiative into a cash-impact analysis.
- In a none priced consequence philosophy (qualitative issues) the value of the initiative will have be important for the (future) society and it will be difficult to do quantitative calculation of the initiative.

4 Circular Economy

The goal for Circular Economy is to keep products and materials in use for as long as possible. End of Life strategies will avoid bringing materials to landfills or burning as End of Life options. This will bring sustainable benefits both for environment, economy and society (e.g. new job creations). The Circular Economy can then lead to new Business Models and Technical innovation, which can lead to more competiveness and allow increased production without more use of natural resources (energy and material). One aspect is also to focus on possible industrial symbiosis where used materials from one industry can be used as input or raw

material for another. A Swedish case study analyses the effects of three strategies underpinning a circular economy: renewable energy, energy efficiency and material efficiency. The case study concludes that by 2030, carbon emissions could be cut by almost 70% if aspects of circular economy policy measures were implemented [13] (Fig. 3).

As an illustration of nations' consumption, there are estimates of how many planet earths that are needed if all other countries had the same consumption level. If Norway set the standard of consumption for all other nations, we would need 2.8 planet earths. The Table 1 shows a limited selection of countries for comparison (2015-figures) [4].

Design for environment is also a demand for a sustainable value chain. Design for environment combined with circular economy solutions do not necessarily imply solutions that will be more expensive. Circular economy starts at the beginning of a product's life. Both the design phase, production processes and the End of Life Strategies have impact on resource use and waste generation throughout a product's lifecycle. Improved product design is one key to facilitating reuse and recycling and contribute to make products that are easier to service and more durable, thus saving precious resources and thereby habitats. Improved production processes will use resources more efficiently, generate less waste, and thus increase the competitiveness for industry [3]. Issues to consider in an environmental design setting can be the following;

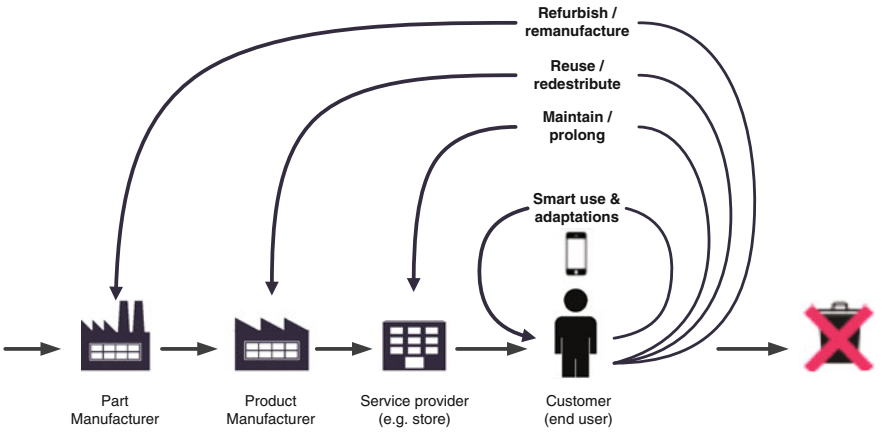


Fig. 3 Framework for products within manufacturing circular value chain; The figure is an elaborated version of the circular economy figure to Ellen MacArthur Foundation/McKinsey [8]

Table 1 Example of Countries Global Footprint in number of earths

Country	Norway	China	USA	UK	India	Kuwait	East-Timor
"No. of Earths"	2.8	1.4	3.9	2.4	0.5	5.1	0.2

- Lifecycle and EoL (End of Life) assessments
- Lifecycle and EoL costs
- Total environment design
- EoL treatment scenarios
- Disassembly structure and Eol process planning

5 An Industrial Example; Finnfjord Smelting Plant (*"Finnfjord Smelteverk"*)

Finnfjord Smelting Plant, in northern Norway, claims to be the world's most energy effective and environment friendly producer of ferrosilicon (100,000 ton/year) and produces micro silica (20,000 ton/year). In addition to this they also produce electric power and algae. The electricity is produced by a steam turbine, which exploits the exhaust heat from the smelter. The steam turbine can produce up to 340 GWh electric power/year.

The algae production project, in cooperation with the University of Tromsø in Northern Norway, concludes a new research project on using exhaust heat and gas for cultivation of algae. For reproduction the algae's also need phosphate. The smelting plant get phosphate from nearby fish farms, for whom phosphate is a biproduct seen as waste or emission. The algae can later be used to salmon food and a dietary supplement.

The reason for farming algae at the smelting plant is that the algae can be fed with CO_2. With use of the algae, Finnfjord Smelting Plant wants to reduce their CO_2 emissions with 50%; from 300,000 tons per year to 150,000 tons per year. With the right results from the research project, the CO_2 emission can be reduced to zero CO_2 emissions. In the future, the smelting plant can both do production of algae and salmon. The CEO of Finnfjord Smelting Plant, considers this as real circular economy [11].

6 Conclusions

Humans are dependent on well-functioning nature and ecosystems. In addition, industrial development needs to pay attention to this and understand that industries and humans have to behave in an environmental and sustainable way. We see the importance of this in global warming and climate change and biodiversity losses. Industry and industry infrastructure need to be developed, building and operating without disturbing or destroying the natural ecosystems. Impact consequences analysis the industrial developments from a sustainable perspective have to be made in new projects; cleansing air and emissions, storing of CO_2, wetlands' CO_2 preserving, reduces danger of floods, water storage etc.

However, an approach like circular economy seems to have significant structural effect on the environmental impact, trying to take care of the lifecycle by maintaining and reusing products in a circular business model.

Maintaining the biodiversity may serve us well in the future, as we may be able to take advantage of species, in manners we are not yet aware of. It is important to take care of our ecosystems and leave enough land/area for other species to maintain sustainable, healthy populations and for humans to let coming generations have possibilities, and benefits from ecosystem services. A lot of today's medicine comes from nature, it can be fungi or substances which are only produced in certain plants of insects.

Other species than the human beings cannot speak up for themselves and thus not defend themselves. Every time we build a new road, a new factory, every time we go bigger and broader and remove habitats other organisms lose their living area. There may be short-term costs on the industry, but the overall benefit is on industry as well as on ecological environment.

Circular economy contributes to a sustainable society and accordingly to sustainable industry. Governmental incentives may be incremental in this early phase of circular economy introduction.

And worth remembering: The ants don't need us, but we need them [2].

References

1. Arlington VA (2015) China's contribution to the Paris climate agreement. In: The Center for Climate and Energy Solutions (C2ES), USA
2. Chivian Eric, Bernstein Aaron (2010) How our health depends on biodiversity. Center for Health and the Global Environment at Harvard Medical School, Boston
3. European Commission (2015) Circular economy implementation of the circular economy action plan. Available via http://ec.europa.eu/environment/circular-economy/index_en.htm
4. Jensen N (2015) Nå er jordas ressurser brukt opp for i år. 13 Aug 2015. Available via http://www.wwf.no/dette_jobber_med/living_planet_report.cfm?47766/I-dag-er-planeten-brukt-opp
5. De Vos JM (2014) Estimating the normal background rate of species extinction. Conserv Biol 29:452–462
6. Ki-moon Ban (2016) The Sustainable Development Goals Report. United Nations, New York
7. Lenart M (2009) An unseen carbon sink. 26 Nov 2009. Available via http://www.nature.com/climate/2009/0912/full/climate.2009.125.html
8. MacArthur E (2015) Growth within: a circular economy vision for a competitive Europe. McKinsey Center for Business and Environment, s.l.
9. Shah A (2014) why is biodiversity important? Who cares? 19 Jan http://www.globalissues.org/article/170/why-is-biodiversity-important-who-cares
10. Sonnemann G et al (2005) Integrated life-cycle and risk assessment for industrial processes. Lewis, Boca Raton, Florida. s.l.
11. Stensvold T (2017) Smelteverket bruker restvarmen til å dyrke alger og utnytter eksosvarme til el-produksjon. 13 March, Available via https://www.tu.no/artikler/smelteverket-bruker-restvarmen-til-a-dyrke-alger-og-utnytter-eksosvarme-til-el-roduksjon/377858?utm_source=newsletter-2017-03-13&utm_medium=email&utm_campaign=newsletteraut

12. Vos M et al (2015) The circular economy. https://www.arcadis.com/media/9/D/3/%7B9D33B0CB-3F9D-4C16-9C74-B763D4BA442C%7DBriefing%20Paper%20-The%20Circular%20Economy_002.pdf
13. Wijkman A (2015) The Guardian Sustainable Business Circular economy could bring 70 percent cut in carbon emissions by 2030. 15 April, Available via https://www.theguardian.com/sustainable-business/2015/apr/15/circular-economy-jobs-climate-carbon-emissions-eu-taxation
14. Wikipedia (2016) Parisavtalen. Available via https://no.wikipedia.org/wiki/Parisavtalen

Lightweight Design of LED Street Lamp Based on Response Surface Method

Mengjia Qi, Lilan Liu, Renfei Ma and Yi Wang

Abstract The junction temperature model of LED street lamp is established by response surface method in this paper. And the accuracy of the model is verified. The key design parameters are obtained through sensitivity analysis. According to the model and the key design parameters, this paper establishes lightweight design mathematical model of LED street lamp. Then an improved particle swarm optimization algorithm is proposed to optimize the design parameters. At last, the LED street lamp structure scheme is verified and analyzed, and the performance and reliability of optimized design are compared and analyzed. It proves the feasibility of method which conducting lightweight design and analysis of LED street lamps.

Keywords LED street lamp · Lightweight design · Response surface method Penalty function

1 Introduction

With the society's increasing demand for energy saving and emission reduction, lightweight design has attracted more and more attention [1]. The reduction of the weight of the LED street lamp can reduce the carbon emission in the production stage and improve the reliability of the LED street lamp. But the traditional design focuses on the performance of LED street lamps, and the research of lightweight design of LED street lamps is obviously insufficient.

Lightweight design of LED lamps has the following aspects: Wang et al. [2] used multi-objective genetic algorithm to optimize fin thickness and fin height of radiator. Considering the weight of the LED street lamp and junction temperature,

M. Qi · L. Liu (✉) · R. Ma
Shanghai Key Laboratory of Intelligent Manufacturing & Robotics,
Shanghai University, Shanghai, China
e-mail: QMJ_SHU@163.com

Y. Wang
School of Business and Management, Plymouth University, Plymouth, UK

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_51

three optimal schemes are obtained: the lowest temperature, the smallest weight, the best combination of the both, and the weight of which can be reduced by 18% at most. Jie et al. [3] used the orthogonal experiment method to optimize LED downlight. To ensure the junction temperature unchanged, the radiator weight reduced by 18%. In this paper, the response surface method is used to establish the junction temperature model of LED street lamps, and the improved particle swarm algorithm is applied to the lightweight design of LED street lamp.

2 LED Street Lamp Model

A LED street lamp is used as an example to study and discuss the application of response surface methodology in the structural design of LED street lamp. The power is 150 W. The exploded figure is shown in Fig. 1.

3 Model and Sensitivity Analysis of LED Street Lamp Junction Temperature

3.1 Summary of Response Surface Method

Response surface methodology is a combination method of mathematical methods and statistical methods. It's a method of obtaining the relation between input (random variables) and output (system response) of complex systems through statistical experiments [4]. The essence is to fit the experimental data and obtain the approximate expression of the system function. Using response surface method to establish mathematical model, the following three problems should be solved: the selection of function fitting form, the extraction method of test sample points, and the evaluation of response surface function [5].

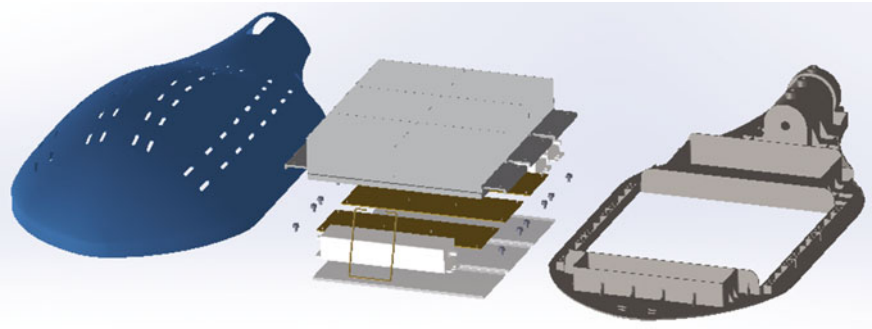


Fig. 1 The exploded figure of LED street lamp

The Selection of Function Fitting

Because the function relationship between the variables and the objective is unknown, the form of the approximation function $\tilde{y}$ must be chosen.

The selection of the response surface function needs to satisfy two conditions. On the one hand, the response surface function should be as simple as possible to meet the requirements. On the other hand, the response surface function should have less undetermined coefficients. The linear and quadratic polynomial is widely used, as follows [6].

$$\tilde{y}(x) = \alpha_0 + \sum_{j=1}^{n} \alpha_j x_j \tag{1}$$

$$\tilde{y}(x) = \alpha_0 + \sum_{j=1}^{n} \alpha_j x_j + \sum_{j=1}^{n} \alpha_{jj} x_j^2 \tag{2}$$

$$\tilde{y}(x) = \alpha_0 + \sum_{j=1}^{n} \alpha_j x_j + \sum_{i=1}^{n} \sum_{j=i}^{n} \alpha_{ij} x_i x_j \tag{3}$$

α_0 is the constant coefficient of the quadratic polynomial, α_j is the primary coefficient of the quadratic polynomial, α_{ij} is the quadratic coefficient of the quadratic polynomial.

The Extraction Method of Test Sample Points

In order to determine the undetermined coefficient of formula (1)–(3), a fixed number of sample point data should be prepared. The fitting accuracy of the response surface function is determined by the distribution of the sample points in the design space. The selection of sample points needs to be followed by some methods, such as full factorial design, fractional factorial design, central compound design, lain hypercube design, uniform design, and orthogonal experimental design. The orthogonal experimental design is used to select the sample points in this paper. It picks out some typical points from a comprehensive experiment based on orthogonality [7]. These typical points have the advantage of uniform dispersion [8].

The Evaluation of Response Surface Function

Because the fitting accuracy of the response surface function has a direct influence on the design results, it is necessary to evaluate the response surface function. Because of the complex structure of LED street lamps, it will increase the cost if the mold is processed to test the real value of the junction temperature of LED street lamps. Therefore, it is reasonable to adopt simulation values instead of real test values during the product design stage. It can provide information support for the optimization design of structure and save research and development time.

$$\varepsilon = \frac{|\text{simulation result - fitting result}|}{\text{simulation result}} \times 100\% \tag{4}$$

3.2 Model of LED Street Lamp Junction Temperature

Because the junction temperature of LED street lamp in traditional method is obtained by software simulation, it is impossible to describe the relationship between the junction temperature of street lamp and the key design parameters. But the junction temperature model based on response surface method can achieve quantitative design, analysis and optimization. The mathematical expression of the junction temperature model is:

$$T_j = f(DDPs) \tag{5}$$

T_j is the junction temperature of LED street lamp, DDPs (Design Dependency Parameters) are the key design parameters, including b (thickness of radiator bottom), d (slot width), N_1 (number of slots), H_t (thickness of fin), H_w (clearance of fin), H_l (height of fin), N_2(number of convection holes), e(radiation coefficient).

According to the formula (3), the model of junction temperature is fitting by the form of quadratic polynomial response surface functions with cross terms.

$$T_j = \alpha_0 + \sum_{j=1}^{n} \alpha_j x_j + \sum_{i=1}^{n} \sum_{j=i}^{n} \alpha_{ij} x_i x_j \tag{6}$$

n is the number of key design parameters, x_i,x_j is the key design parameters, such as b (thickness of radiator bottom), d (slot width), N_1 (number of slots), H_t (thickness of fin), H_w (clearance of fin), H_l (height of fin), N_2 (number of convection holes), e (radiation coefficient), α_0, α_j, α_{ij} is undetermined coefficient.

3.3 Modeling of LED Street Lamp Junction Temperature Based on Response Surface Method

Experimental Design

The junction temperature is chosen as the experimental object, and the key design parameters are chosen as the experimental factors. The range of experimental factors: b = [2–4]; d = [1–3]; N_1 = [1–3]; H_t = [1,1.75,2]; H_w = [1–3]; H_l = [36,41,46]; N_2 = [64,72,80]; e = [0.90,0.98].

The experimental table was designed by orthogonal experiment. Thermal simulation of the experimental scheme is performed by using FloEFD. The experimental scheme and results are shown in Table 1.

Table 1 Orthogonal experimental scheme and results

No.	Experimental factors								Simulation result (°C)
	b (mm)	d (mm)	N_1	H_t (mm)	H_w (mm)	H_l (mm)	N_2	*e*	
1	2	1	1	1	1	36	64	0.98	78.88
2	2	2	2	1.75	2	41	72	0.98	78.86
3	2	3	3	2.5	3	46	80	0.98	77.47
4	3	1	1	1.75	2	46	80	0.98	76.80
5	3	2	2	2.5	3	36	64	0.98	78.13
6	3	3	3	1	1	41	72	0.98	75.64
7	4	1	2	1	3	41	80	0.98	77.58
8	4	2	3	1.75	1	46	64	0.98	78.11
9	4	3	1	2.5	2	36	72	0.98	76.91
10	2	1	3	2.5	2	41	64	0.90	77.30
11	2	2	1	1	3	46	72	0.90	78.73
12	2	3	2	1.75	1	36	80	0.90	79.30
13	3	1	2	2.5	1	46	72	0.90	78.13
14	3	2	3	1	2	36	80	0.90	79.70
15	3	3	1	1.75	3	41	64	0.90	78.33
16	4	1	3	1.75	3	36	72	0.90	78.89
17	4	2	1	2.5	1	41	80	0.90	77.18
18	4	3	2	1	2	46	64	0.90	78.77

The response surface fitting function is obtained according to the orthogonal experimental table, as shown in formula (7).

$$\begin{aligned} T_j = {} & 2.5116N_2 + 0.0931b \cdot H_l - 0.0652b \cdot N_2 \\ & -0.3095d \cdot N_1 + 0.0682d \cdot H_l - 0.0361d \cdot N_2 \\ & +0.7748N_1 \cdot H_w + 0.1817N_1 \cdot H_l - 0.1205N_1 \cdot N_2 \\ & -0.0090N_1 \cdot e + 0.1337H_t \cdot H_l - 0.0870H_t \cdot N_2 \\ & -0.0750H_w \cdot H_l + 0.0367H_w \cdot N_2 + 0.0019H_l^2 \\ & -0.0157H_l \cdot N_2 - 0.0073N_2^2 - 0.2411N_2 \cdot e \end{aligned} \tag{7}$$

The Evaluation of Response Surface Function

Thermal simulation of the ten randomly generated design parameters is performed by using FloEFD. The response surface function is evaluated according to the results by formula (4), as shown in Table 2.

It can be seen from Table 2 that the relative error is about 0.5%, the fitting results are in good agreement with the simulation result. The fitting evaluation results are shown in Fig. 2.

Table 2 Comparison of fitting function and thermal simulation

No.	b (mm)	d (mm)	N_1	H_t (mm)	H_w (mm)	H_l (mm)	N_2	e	Fitting result (°C)	Simulation result (°C)	Relative error (%)
1	4	1	1	1	1	36	72	0.98	81.64	79.73	2.4
2	4	2	1	1.75	2	41	80	0.98	80.86	80.32	0.68
3	3	3	1	2.5	3	46	64	0.98	81.71	81.07	0.46
4	3	1	2	1.75	2	36	72	0.98	79.71	79.57	0.18
5	2	2	2	2.5	3	36	64	0.98	81.39	81.16	0.28
6	2	3	2	1	1	41	64	0.9	82.97	83.44	0.57
7	4	1	1	1	3	46	72	0.9	78.03	77.71	0.42
8	3	2	3	1.75	1	41	80	0.9	79.09	81.29	2.71
9	4	3	3	2.5	2	36	80	0.9	79.65	79.28	0.47
10	2	1	3	2.5	2	41	64	0.9	80.28	79.89	0.48

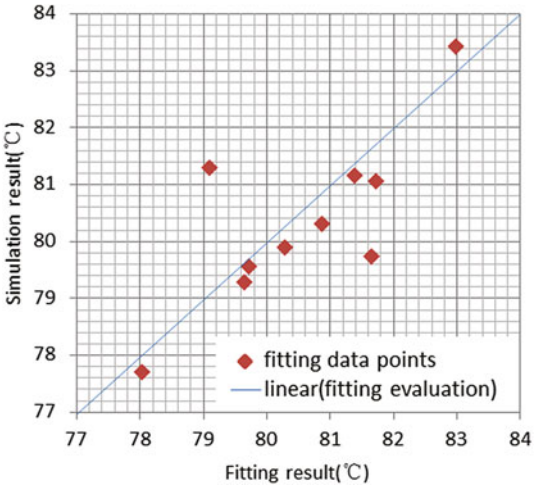


Fig. 2 Fitting evaluation results

3.4 Sensitivity Analysis

The sensitivity of the junction temperature of the LED street lamp to random variables $X = [X_1, X_2, \cdots, X_n]$ is

$$\frac{\partial R}{\partial X_i} = \frac{\partial}{\partial X_i}\left(\alpha_0 + \sum_j^n \alpha_j x_j + \sum_i^n \sum_{j=i}^n \alpha_{ij} x_i x_j\right) \tag{8}$$

The response surface function of junction temperature function is substituted into formula (8), and the sensitivity formula of each design parameter is obtained. The orthogonal experimental design scheme 6 is taken as the initial scheme. The relevant parameters are substituted into the formula, and the sensitivity is

$$\left[\frac{\partial T_j}{\partial b} \quad \frac{\partial T_j}{\partial d} \quad \frac{\partial T_j}{\partial N_1} \quad \frac{\partial T_j}{\partial H_t} \quad \frac{\partial T_j}{\partial H_w} \quad \frac{\partial T_j}{\partial H_l} \quad \frac{\partial T_j}{\partial N_2} \quad \frac{\partial T_j}{\partial e}\right] =$$
$$[-0.88 \quad -0.73 \quad -1.39 \quad -0.78 \quad 1.89 \quad 0.11 \quad -0.14 \quad -17.39]$$

As we can see, the sensitivity of the design parameters (b, d, N_1, H_t, N_2, e) is negative. The greater the value, the more sensitive the junction temperature is to the design parameters. Sort by degree of influence of junction temperature: e(radiation coefficient) > N_1(number of slots) > b(thickness of radiator bottom) > H_t(thickness of fin) > d(slot width) > N_2(number of convection holes). The sensitivity of other parameters is positive. Sort by degree of influence of junction temperature: H_w (clearance of fin) > H_l(height of fin).

4 Lightweight Design of LED Street Lamp

4.1 Lightweight Design Mathematical Model of LED Street Lamp

The junction temperature and the range of key design parameters is used as constraint and volume is the optimization object. The mathematical model is

$$\begin{cases} \min f(x) \\ s.t. T_j(X) \geq T_0 \\ g_u(X) \geq 0, u = 1, 2, 3, \cdots, m \end{cases} \tag{9}$$

$T_j(X)$ is the junction temperature of LED street lamp, T_0 is the junction temperature of initial scheme, $f(X)$ is the volume of LED street lamp, $g_u(X)$ is other constraints.

4.2 An Improved Particle Swarm Optimization Algorithm

The inertia weight coefficient ω of particle swarm optimization algorithm is an important parameter. A larger ω is beneficial to enhance global searching ability, but a smaller ω is beneficial to enhance local search ability. In order to balance the global search ability and local search ability, a self-adaption weight is used to improve the inertia weight coefficient. The inertia weight coefficient is shown in formula (10).

$$\omega = \begin{cases} \omega_{\min} - \frac{(\omega_{\max} - \omega_{\min}) \times (f - f_{avg})}{(f_{avg} - f_{\min})}, f \leq f_{avg} \\ \omega_{\max}, f > f_{avg} \end{cases} \tag{10}$$

$\omega_{\max}$, $\omega_{\min}$ are the maximum and minimum values of the inertia weight, f is the current object function value of the particle, f_{avg}, $f_{\min}$ are the average object value and minimum object value of all current particles. In this paper, the value of $\omega_{\max}$ is 0.9, the value of $\omega_{\min}$ is 0.4.

Multistage penalty function is used to solve the constrained optimization problem of particle swarm optimization in this paper. It concerts a constrained problem to an unconstrained one. The objective function constructed by penalty function is

$$F(x) = f(x) + h(k)H(x),\ x \in S \subset R^n \tag{11}$$

$f(x)$ is the original objective function, $h(k)$ is penalty factor of penalty function, k is the number of iterations of PSO, $H(x)$ is multistage penalty function, generally defined as [9]:

$$H(x) = \sum_{i=1}^{m} \theta(q_i(x)) q_i(x)^{\gamma(q_i(x))} \tag{12}$$

$$q_i(x) = \max\{0, g_i(x)\}, i = 1, 2, 3, \cdots, m \tag{13}$$

m is the number of constraints, $q_i(x)$ is the degree violating the constraint, $g_i(x)$ are constraint functions, $\theta(q_i(x))$ is multistage distribution function, $\theta(q_i(x))$ is the progression of a penalty function. The improved PSO is used for lightweight design of LED street lamp. The optimal design process is shown in Fig. 3.

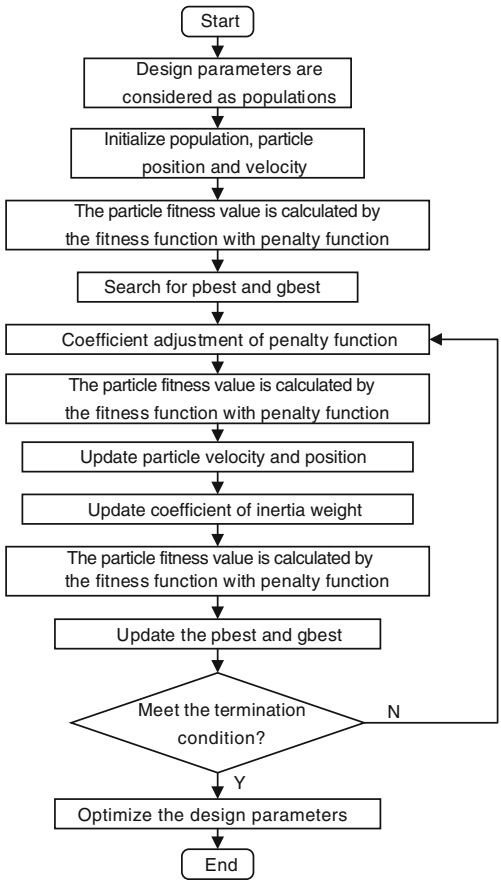


Fig. 3 An improved particle swarm optimization algorithm

Table 3 The comparison of optimization scheme

	Junction temperature (°C)	Volume (mm^3)
Initial scheme	75.64	1914523.4
Optimized scheme	71.99	1490247.2

4.3 Comparative Analysis of Optimization Results

After optimization and rounding the parameters, the key design parameters are obtained.

$$[b, d, N_1, H_t, H_w, H_l, N_2, e] = [3.8, 2.7, 3, 1, 1.8, 36, 78, 0.95]$$

The thermal simulation of the optimized key design parameters is made. Compare the optimized scheme with the former one, and the results are shown in Table 3.

It can be seen that the junction temperature of the LED street lamp is reduced by 4.8%, and the volume is reduced by 22.2%. It means that after lightweight design, the quality of LED street lamp is reduced by 22.2%.

5 Conclusions

Analysis and research on lightweight design of LED street lamp are carried out in this paper. The key design parameters of the LED street lamp are taken as the design variables, and the volume is taken as the design objective. Finally, the quality of the LED street lamp is reduced by 22%, and the performance of the LED street lamp is improved by reducing junction temperature. The feasibility of the proposed method is proved. Meanwhile, the method is suitable for other LED lamps. Applying this method to the preliminary design of the LED lamps can reduce the cost of design and production effectively, and improve the performance of the products.

Acknowledgements The authors would like to express appreciations to mentor in Shanghai University for its valuable comments and other helps. Thanks for the funding of Shanghai Science and Technology Committee of China and Shanghai Economy and Informationization Committee of China. The program numbers are No. 15521103502 and No. 15X I-1-26.

References

1. Tsunenari S, Oya T (2016) Method for evaluating mechanical characteristics of biological material for bio-inspired lightweight design. Comput Aided Des Appl 13(4):503–510
2. Wang L, Wang G, Fan B, Li J (2016) The heat simulation and optimization design for the high power LED tunnel lighting. Mach Des Manuf 16:19–21+25

3. Jie C, Xiaohong L, Zhong C, Meizhen Z, Yashu C (2014) Research of orthoaonal experimental optimum design of LED downlight radiator. Semicond Optoelectron 03:480–483
4. Su Y, Wang X, Zhang Z-s (2011) Response surface methods for reliability analysis and its engineering application. J Eng Geol 09:381–384
5. Sui Y, Yu H (2014) Improvement and application on response surface methodology in shape optimization. J Beijing Inst Petro-chem Technol 01:31–36+64
6. Chennupati S, Potumarthi R, Rao MG, Manga PL, Sridevi M, Jetty A (2009) Multiple responses optimization and modeling of lipase production by Rhodotorula mucilaginosa MTCC-8737 using response surface methodology. Appl Biochem Biotechnol 159(2):317–329
7. Parra R, Aldred D, Magan N (2005) Medium optimization for the production of the secondary metabolite squalestatin S1 by a Phoma sp. combining orthogonal design and response surface methodology. Enzyme Microb Technol 37(7):704–711
8. Kyriazis GA (2008) An orthogonal design of experiments for accurately estimating harmonics of periodic signals at low frequencies. IEEE Lat Am Trans 6(1)
9. Z Guoying, W Yijuan (2014) Multi-constraint optimization algorithm based on multistage punish function and particle swarm optimization. J Beijing Inst Petro-chem Technol 16:30–32

Industry 4.0 and Cyber Physical Systems in a Norwegian Industrial Context

Ragnhild J. Eleftheriadis and Odd Myklebust

Abstract Looking at the Norwegian industry, it generally consists of small and medium size enterprises. Many of them are suppliers to the oil and gas, fishery and maritime industries. These industries have all been strong contributors to the Norwegian growth. However, many of these suppliers are very vulnerable for market changes and the dependencies of oil and gas prices. In 2016, the EU's Digital Agenda Scoreboard and Societal Index put Norway among the front when it comes to digitalization (Digital Single Market, Digital Economy and Society in Progress by country, Statistics, Norway/Europe. European Commission https://ec.europa.eu/digital-single-market/en/scoreboard/norway, 2017 [1]). Some important public sectors were highlighted, but also Industry and the society in general have been subject to digitalization. Online banking, public authorities, the Norwegian tax system; register of companies and large health studies has been digitalized. One main reasons is mentioned that Norway has one of the most organized labour markets in Europe, this for both employers and employees. The Norwegian labour policy is firmly based on a tripartite co-operation between the government, trade unions and enterprise federations, with a low hierarchical system between employers and employees, which is mentioned as "The Norwegian Model". Many Norwegian companies have developed lean management methods, in public sector as well as industrial manufacturing. Methods with highly qualified and self-depended labour force. How will existing methods as Lean in a high cost country as Norway be develop with implementation of Industry 4.0 and Cyber Physical Systems. We have looked at several industry branches in one county in Norway and this paper present some reflections from this Industry 4.0 survey. This to show the challenges the industry are facing with strategies, implementation and adaption of flexible and smart production interpreted for implementation of Cyber Physical Systems (CPS) in a Norwegian context.

R. J. Eleftheriadis (✉) · O. Myklebust
SINTEF Raufoss Manufacturing AS, S.P. Andersens v.5, 7031 Trondheim, Norway
e-mail: ragnhild.eleftheriadis@sintef.no

O. Myklebust
e-mail: Odd.Myklebust@sintef.no

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_52

Keywords Industry 4.0 · Cyber physical systems · Digitalization Management

1 Introduction

The macro economical model and the labour policy in Norway have been regulated by a tripartite co-operation between the government, trade unions and enterprise federations for decades [2, 3]. These regulations have led to democratic rights, balanced labour regulations and an accomplice liability for workers. The employers as well as the employees together with the unions have made this possible with a wide and smooth cooperation, and they have given the Norwegian companies a working organization based on autonomy and learning, which may have increased innovation and global competition potential [4, 5]. Due to early adoption of Safety, Health and Environmental (SHE) regulations often initiated by the workers themselves and with help from trade union, Norwegian companies have adapted and implemented different types of quality programs like TQM, ISO9000 and Lean.

Lean management and lean manufacturing are one of the most used methods with a strong basis in many industrial companies [6]. The demand for continuous improvement, employee involvement and standardization of tasks are depended on highly qualified and self-dependent workers. In the view of a newly developed industrial strategy for Norway [7], it is a need for a strong manufacturing capability, to be able to digitalize the Norwegian industry. The paper will illustrate some of the challenges and possibilities the Norwegian industry is facing, with a focus on the land-based industry. There are especially two specific areas of interest; the implementation phases of Industry 4.0 and strategic leadership. However, during an Industry 4.0 survey, we have spotted some interesting findings and problem areas. This is also mention earlier by a German Industry 4.0 survey [8] and we can see some similarities and differences between cultural aspects. In the discussion later in this paper we wanted to explain and compare some of this finding further.

2 Relevant Literature and Industry 4.0 Survey's

2.1 The "Norwegian Model" and Management Role

The implementation from Lean to Industry 4.0 and further to CPS is for most companies difficult to face. Many companies give feedback on that they are missing a strategic tool, a direction for how to start and where to go on from typical manual based quality control systems to a more digitalized production and quality system. This request a need for a strong industrial governmental directions, strategic tools, skilled workers and smoot R&D development in industrial companies.

This for making it possible for the management to draw a roadmap and a strategic direction for where production in the new digital era develops. Each company needs to create their own strategic impact to be productive in the new development. Work organizations are not affected by the macro-political order of society alone, but also more specific initiatives that are taken; to promote organization development at the workplace level says Gustavsen [3]. When he talks about the "the Scandinavian Model" he says that the number of bi- and tripartite efforts to promote learning-oriented forms of work organizations. With the "Scandinavian Model", he mean that from his view the Scandinavian countries have own and similar characteristic for working development. Which lies in the cultural aspect of values and standard of society [9, 10]. This is partly agreed by Levin [2], but he argues that the governmental cooperation between the labour and workers in Sweden is lower than in Norway, and that Denmark has moved towards the common EU platform. However, when it comes to regulations in the working environment and continuous improvement Scandinavian countries are still quite similar [2, 11].

2.2 *Lean Manufacturing Versus Industry 4.0, CPS and Zero Defect Manufacturing*

The kernel of lean manufacturing is to create a streamlined flow of processes to create the finished products at the required pace of customers and with no or less as possible waste [12]. Lean as a methodology is described as an approach comprising industrial practices and identifying value streams, for adding value to customer and pull throughout the organization [6]. Some of this focuses were changed when Taichii Ohno's initiatives were introduced to Toyota Motor Company with the Toyota Production System (TPS) [13] with the two different scenarios; Just in time and automation.

To understand the differences between the Lean concept and Industry 4.0 perspective, it all lies in the manual processes. While Lean as a cost reducing quality concept in the second and third stage of the industrial evolution. Industry 4.0 and CPS have reached the fourth stage, this has earlier been visualized by a curve showing how manual and automated process steps have adopted the CPS and Industry 4.0 development [14] and the figure from Industry 4.0 illustrate the historical perspective in a nice way. Industry 4.0, CPS and Zero Defect Manufacturing (ZDM), focus also on how to do the manufacturing processes without defect, even if the cost of work will be more expensive at first [15–18] (Fig. 1).

The IFaCOM Framework is a part of the Zero Defect Manufacturing (ZDM) Framework (Fig. 2) with all the loops online and self-adaptive feedback system loops [17]. While the novel multi model method for CPS, based on prediction of manufacturing lead-time for handling and realizing use of statistical learning methods and reactive/proactive operation modes for simulation is show in Fig. 3 [19].

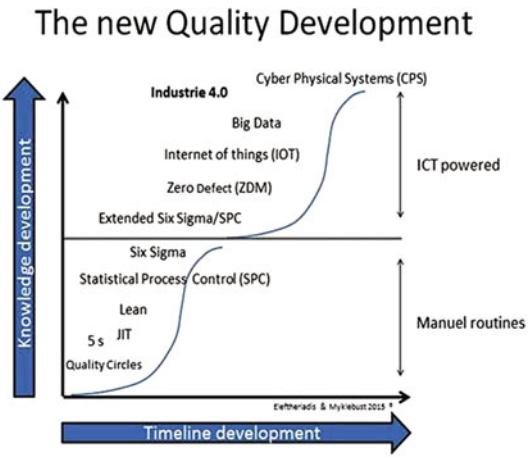


Fig. 1 Quality building blocks of a digitalized development [14]

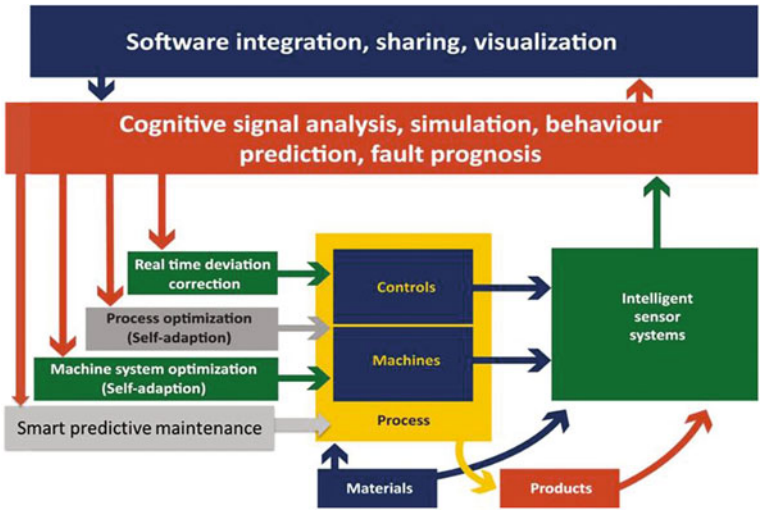


Fig. 2 The IFaCOM closed loop framework [17]

2.3 *Lean Manufacturing Versus Industry 4.0, CPS and Zero Defect Manufacturing*

As shown on the above figures ZDM and CPS has their own frameworks, however, one of the key aspects of Industry 4.0 is the RAMI, The Reference Architectural Model Industry 4.0. This architectural model provides an interesting view point of Cyber Physical Systems and combines the relations between product-lifecycle and the assets/factory lifecycle in a "Lifecycle Value Stream" [20].

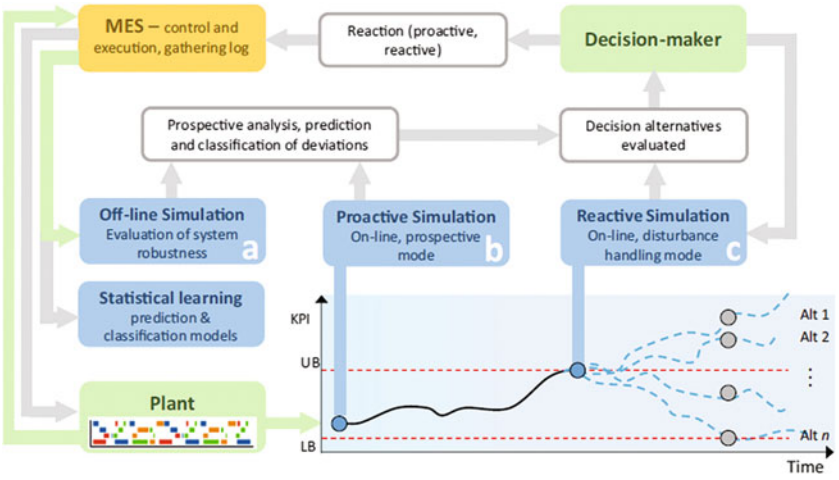


Fig. 3 Multi model for CPS, based on prediction and simulation of lead-time [19]

This integrate the CPS model as a Cyber Physical Production System (CPPS). The factory lifecycle axes both the physical and digital ones are considered, and the enterprise modelling is the key factor. This creates a unique data and information flow that will first link all stakeholders in the manufacturing enterprise domain but also integrate the manufacturing execution system (MES). The MES can reschedule the jobs to mitigate the impact of breakdown and lead-time [19, 21].

With more advanced analytics and big data environment, machines are equipped to be more self-aware and self-maintained. Such machines can assess their own condition and degradation and utilize data from other machines to avoid potential maintenance issues [22]. In the manufacturing environment, the CPSs comprise smart storage and flow systems and production facilities capable of autonomously exchanging information, triggering actions and controlling each other independently. CPS and CPPS is the description of the physical assets connected to the virtual worlds, the methodology will create new frameworks for state of the art production systems, with integrated CPS and IoT technology [19]. The approach will open the opportunity of developing improved self-learning equipment, with novel software and hardware, which will allow their output optimization [16, 23].

The Cyber-Physical Systems (CPS) are embedded computers and networks that monitor and control the physical processes, usually with feedback loops where physical processes affect computations done in the control programmes [14, 18]. Among the objectives of CPS as an approach are to get zero defects (ZDM) in a production environment, to eliminate waste and scrap reduction, lower production costs, shorter production times, higher productivity and competitiveness, and last but not least this also influences on the material, resource and energy efficiency. These goals can be found in the Lean concepts also, and both philosophies highlighted that they bring significant improvement to the work environment for the

manufacturing industry. However, the development and implementation of suitable ZDM tools as solutions for the upstream generation and downstream propagation of production defects makes CPS an innovative tool [14].

3 Introduction to a Manufacturing Survey on Industry 4.0

In a Norwegian survey, we asked one hundred companies to see how far the implementation of Industry 4.0 and adaption of flexible and smart production interpreted by typical manufacturer. Nine different manufacturing branches and some technology providers got the survey and invitation to a workshop where presentation of the results and discussion of the latest services and products in CPS, Industry 4.0 and Artificial Intelligent (AI) was provided.

The survey was concentrated on a limited geographical area and with only production manufacturers. They answered around thirty questions where the concentration was around management, competence and relation to the Industry 4.0. The largest groups of manufacturer were companies in mass production, customised production and small series, most of them was producing electrical components, machine and metal parts. There were also a large group making composite and plastic materials, which was mostly suppliers in oil and marine operation.

All the managers make a self-assessment of their own production processes in relation to Industry 4.0, the innovation potential and strategic implementation of a digital roadmap for their own company. More than 34% of the answers was telling that they have never written down any kind of documentation of Industry 4.0, either create new business models for their company. Many of them was well suited in the operation faces but failed when it comes to the implementation of own innovation potential. They could see the potential but was not able to implement this into a product. Innovation faces is difficult to implement and competences in product design and production planning is some few focus areas they need for implementation faces. Many of the companies had very skilled workers, however, most of them answered that that they felt a lack of competences and wanted more education into the specific field like “Big Data, Statistics, AI and IOT”. We also asked their interest of bringing production back to Norway if they have had more competence and digital skills internally. Here we got surprisingly 80% saying that with the use of Industry 4.0, the automation potential for production could be digitalised and produced in Norway, even if this lead to less work for a short time period.

In a similar German survey for the Rhine-Neckar region [8], they found much of the same challenges especially in the area of production complexity and creation of business-models. However, the investment cost and cyber security did not get the same focus by the Norwegian managers and the IOT infrastructure, which for Norwegian manufacturing companies this quite good, even for a spread country as Norway and one main reason could be heavy infrastructure work from ICT companies and government side.

4 Can "the Norwegian Model" Be an Effective Implementation of Digital Solutions in a Global Market?

The earlier mentioned new industrial strategy from the Norwegian government have looked at different scenarios for how Norway need to change from oil dependencies to more sustainable oriented industries combined with the digital market. Digital transformation of society it selves creates many new needs. A good way to respond to these needs is to put emphasis on giving the workforce appropriate and adaptable competences through education and training.

Frey and Osborne reported in 2013 on the computerization in the labour marked and they have looked at how computerization influences on job creation in several countries including Norway [5]. Memberships in the trade unions, low hieratical structure and wide involvement in organisational question can be a key to globalisation. In some part of the world, public sector traditionally have been seen as a place of waste and inefficiency. From a Lean perspective, this is how a need for improvement appears, but in fact, the public sectors in the Nordic countries is probably an asset for competitiveness and innovation [4]. In a digital marked, this might be the structure you need for make a fast access to a global industrial market, where digital skill, highly educated population and a smooth working market is essential for survival.

5 Conclusions

This paper has look upon the aspects of Lean, Industry 4.0 and Cyber Physical System in manufacturing in a Norwegian context. We have seen that implementation of Industry 4.0 has certain barriers particular in small companies. Most of these barriers lies in the educational and managerial area, where the managers often focus on the operational day-to-day activities. Response on long-term strategies and knowledge of new technology is therefore often an area for improvement on managerial skills. An implemented Lean concept could therefore be a barrier for the management, since Industry 4.0, ZDM and CPS is operating in different development than Lean. Industry 4.0, ZDM and CPS is therefore in managerial perspective more proactive; you need independent workers to response on operational corrections and let the manager lead the organisation.

From our survey, we also know that lack of skills and education in general will make a huge difference between small and medium sized organisations. Medium-sized organisations is more equipped to do own R&D, and have gravity to employ skilled workers. It is still a bit early to conclude and literature stated that there are certain ways of scaling our indicators in the Norwegian society that bring digitalization forward, some more surveys and research is necessary for finding this answers.

References

1. Digital Single Market, Digital Economy and Society (2017) Progress by country, Statistics, Norway/Europe. European Commission. https://ec.europa.eu/digital-single-market/en/scoreboard/norway
2. Levin M (2012) 4, Den norske Arbeidslivsmodellen., Magma - www.magma.no/den-norske-arbeidslivsmodellen, Vol. 4, pp. 20–23
3. Gustavsen B (2007) Work organization and «the Scandinavian Model». Econ Ind Democracy 28:650–671
4. Schubert CB, Martens H (2005) The Nordic model: a recipe for European success? European Policy Center, Nordic Council of Ministers
5. Pajarinen M, Rouvinen P, Ekeland A (2016) Computerization and the future of jobs in Norway. http://nettsteder.regjeringen.no/fremtidensskole/files/2014/05/Computerization-and-the-Future-of-Jobs-in-Norway.pdf
6. Womack James P, Jones Daniel T, Roos Daniel (1990) The machine that changed the world. Rawson Associated Scribner, New York
7. Den Norske Regjeringen (2017) Regieringen.no Stortingsmelding no27. https://www.regjeringen.no/no/dokumenter/meld.-st.-27-20162017/id2546209/sec1
8. Jäger J et al (2016) Advanced complexity management strategic recommendations of handling the "Industrie 4.0" complexity for small and medium enterprises. In: 49th CIRP Conference on manufacturing systems, CIRP-CMS-2016, Procedia CIRP, Science Direct, Elsevier
9. Grenness Tor (2012) På jakt etter en norsk ledelsesmodell. Magma 04:51–59
10. Hofstede G (2001) Culture's consecvenses, comparing values, behaviors, institutions and organizations across nations
11. Ingvaldsen JA, Rolfsen M, Finsrud HD (2012) Lean organisering i norsk arbeidsliv: slutten på medvirkning? www.magma.no, 4:42–50
12. Shah R, Ward PT (2007) Defining and developing measures of lean production. J Oper Manage 25(4):785–805
13. Ohno T (1988) The Toyota production system: beyond large scale production. 1988. Productivity Press, Portland, Oregon
14. Eleftheriadis RJ, Myklebust O (2016) A quality pathway to digitalization in manufacturing thru zero defect manufacturing practises. Atlantis Press, p 5. http://www.atlantis-press.com/proceedings/iwama-16/25862244
15. Kagermann H, Wahlster W, Helbig J (2013) Securing the future of German manufacturing industry. Recommendations for implementing the strategic initiative INDUSTRIE 4.0. Final report of the industrie 4.0 working group. Forschungsunion and Acatech, s.l.
16. Lee J, Bagheri B, Kao H-A (2015) A cyber-physical systems architecture for Industry 4.0 based manufacturing systems. Science Direct. Manuf Lett 3:18–23
17. Myklebust O (2015) IFaCOM. IFaCOM Final Summary report http://cordis.europa.eu/project/rcn/101390_en.html. European Commission, R & D Department, Trondheim, Norway
18. Monostori L et al (2016) Cyber-physical systems in manufacturing. CIRP Ann Manuf Technol 65(2016):621–641
19. Pfeiffer A et al (2016) Manufacturing lead time estimation with the combination of simulation and statistical learning methods. Procedia CIRP 41:75–80
20. VDI/VDE ZVEI SG2 (2015) Reference Architecture Model Industrie 4.0, Umsetzungsstrategie Industrie 4.0 – Ergebnisbericht. VDI/ VDE Association and Technical Committee 7.21 "Industrie 4.0", Berlin
21. Lucke D, Constantinescu C, Westkämper E (2008) Smart factory—a step towards the next generation of manufacturing. Manufacturing systems and technologies for the new frontier. Springer, London

22. Lee J, Kao H-A, Yang S (2014) Service innovation and smart analytics for Industry 4.0 and big data environment. Product services systems and value creation. In: Proceedings of the 6th CIRP conference on industrial product-service systems, pp 3–8
23. Lee EA (2008) Chyber physical system; design challenges. Technical Report No. UCB/EECS-2008-8, University of California at Berkeley, San Fransico, USA

Mechanical Analysis of a Customized Hand Orthosis Based on 3D Printing

Jiawei Wu, Cuilian Zhao, Yexiao Liu and Shuangchi Ma

Abstract A clinical hand orthosis is used to relieve the finger spasm of stroke patients. Generally, there are three dimension size of the hand orthosis for different patients, but it can not achieve personalized precision treatment. To handle these problems, 3D printing technologies were used to make a customized hand orthosis. In order to make it affordable in clinical application and individual design, this paper provided a mechanical analysis model. First a middle finger model of stroke patients was built, then based on the three-point pressure-correction method, its mechanical properties were tested and its bending strength was calculated by finite element analysis model.

Keywords 3D printing · Hand orthosis · Customized · Mechanical analysis

1 Introduction

In China, stroke is the leading cause of death and disability in the aged. After a stroke, 80% of patients suffered from acute upper limb syndrome, and only 30% could cure. It's difficult to open their hand for functional grasp, seriously affected the quality of patients' daily life, so the hand function rehabilitation of stroke patients is the focus of treatment [1].

It has been confirmed that active extension is good for the rehabilitation of patients with hand spasm, such as bao [2] found that the treatment with hand orthosis was more useful for relieving the pain of finger spasms and promoting the recovery. Therefore, in the early clinical rehabilitation, the use of the hand orthosis has significant in the hand function rehabilitation of patients with high muscle tone.

Wood or injection molding is commonly used for the production of hand orthosis. For example, Ellen Scherling [3] designed three different molds to make

J. Wu · C. Zhao (✉) · Y. Liu · S. Ma
Shanghai Key Laboratory of Intelligent Manufacturing and Robotics, Shanghai University, Jing'an District, Shanghai, China
e-mail: clzhao@mail.shu.edu.cn

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_53

the hand orthosis. These hand orthoses may have effect in rehabilitation, but they can not achieve the accurate treatment, and there are many shortcomings, such as fixed problems, higher cost and so on.

3D printing technologies, that is additive manufacturing (AM) technologies, get wildly used in the orthopedics and rehabilitation fields, because of its low cost and the ability to process complex surfaces. According to above problems, 3D printing technologies can be used to customize the rehabilitation devices for the patient's own. Gabriele Baronio [4] used 3D scanning technologies and 3D printing skills in making personalized hand orthosis.

The hand orthosis, as an fingerboard for rehabilitation, must be mechanical analyzed before clinical application, and the personalized design also requires the mechanical analysis model. Therefore, this article used 3D printing technologies to make personalized hand orthosis, then built the middle finger model of stroke patients with hand orthosis, finally based on the three-point pressure-correction method, making mechanical property tests and the calculation of the bending strength by the finite element analysis model.

2 Customized Hand Orthosis

2.1 3D Model Acquisition of Hand Orthosis

The purpose of this phase explains the critical phase of 3D Model acquisition.

(i) Use handheld scanner (Creaform 3D scanner) to scan healthy hand for a total scanning time of approximately 3 minutes as shown in the Fig. 1a. The hand remains stationary, or the palm and the hand back can't align. After scanning, export the entire hand model for STL format files.

(ii) Import the above model into the reverse modeling software (geomagic studio), get the sick hand model by the mirror operation, cut out the palm of a closed curve, select the outside face of palm and set offset, then do some repair operations, such as removing small protruding parts. Finally, carry out surface modeling and export the hand orthosis model for x-t format files.

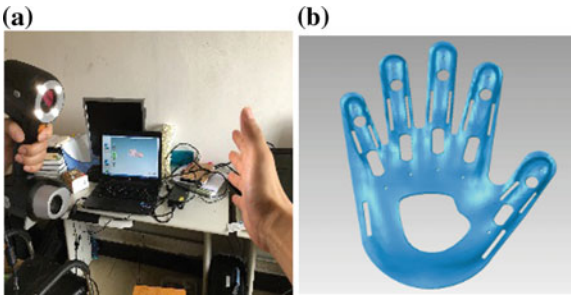


Fig. 1 The 3D model acquisition

(iii) Use computer aided design software (UG) to open the above x-t format file, then design fixed holes and air holes, finally get the hand orthosis model as shown in the Fig. 1b. Then export it of STL format for 3D printing.

2.2 3D Printing Process

This paper used the YW3P-4 3D printer, which is a desktop 3D printer with FDM technology. The progress of 3D printing includes following steps.

Firstly, import the hand orthosis model into 3D printer software (Cura). Secondly, set the palm up and placed it in the printable area, then set the slice and support, then print it. Finally, remove the support structure and polish the printed hand orthosis for sharp processing.

3 Rehabilitation Therapy of the Clinical Hand Spasm

3.1 The Physiological Mechanism of Hand Spasm

Spasticity, as a common symptom of patients with stroke, is crucial for the rehabilitation. Spasticity is a motor disorder characterized by a velocity dependent increase in tonic stretch reflexes with exaggerated tendon jerks, resulting from hyperexcitability of the stretch reflex, as one component of the upper motor neuron syndrome [5]. So spasticity could cause the corresponding limb deformity or abnormal movement pattern, for example typical hand postures include buckling wrist, fist and thumb flexion adduction as shown in the Fig. 2.

The physiological mechanism of hand spasm mainly contains tonic stretch reflex [6]. Stretch reflex refers skeletal muscles under the external force cause reflex of the same muscle, including tendon reflex and muscle tension. Clinical studies shown

Fig. 2 Typical hand posture of stroke patients

that low muscle tone happened during the early paralysis of acute hemiplegia, while muscle tone improved and tendon reflex exaggerated after a few days or weeks. After the stroke, tendon reflex became hyperexcitability and muscle tone increased, and patient's fingers was bended and adduction, these involuntary behaviors obstructed the extension of the fingers.

Thus this paper designed the hand orthosis for helping patients extend there fingers in their daily life, and its rehabilitation therapy is shown in the next phase.

3.2 Rehabilitation Therapy of the Hand Orthosis

The clinical hand orthosis is a kind of external support device for relieving the muscle tone. This hand orthosis refers to physical contact with the patient, then applies force on the finger joint, finally it will change the abnormal balance formed by the internal and external force in patients' body.

Instructions of wearing the hand orthosis: Firstly, stretch the finger of a stroke patient. Secondly, fit the patient's palm and fingers with the hand orthosis. Finally, use fixing bands through the side hole of the finger and palm to fix the patient's finger with the hand orthosis.

Because of hand spasm, patients want to bend their fingers together, then the fixed bands in the joints of fingers exert reaction force to force patients to stretch fingers. This therapy is called the three-point pressure-correction method.

The hand contains the thumb, index finger, middle finger, ring finger and little finger. Middle finger is the most easy to break because of its large span, so build a three joint middle finger model as shown in the Fig. 3, patients' middle finger produce force F to the fixed band, then fingertip and root exert force to the hand orthosis are F_1 and F_2, and then the reaction force of these forces apply to the finger, F_1 and F_2 support the finger and F forces the finger to stretch.

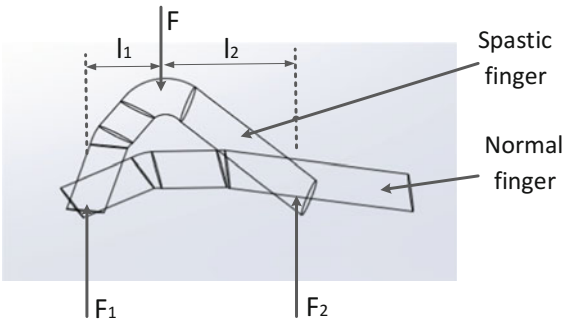


Fig. 3 Middle finger model

4 Mechanical Analysis of the Hand Orthosis

On the basis of the three-point pressure-correction method, the mechanical property test of the hand orthosis is carried out, and the finite element analysis model is established to analyze whether the hand orthosis could be applied to clinical rehabilitation treatment.

4.1 Mechanical Property Test

Before the clinical application of the hand orthosis, we need get maximum bending force of it. Thus after carrying out the bending tests of middle fingers of hand orthoses made of common materials including acrylonitrile butadiene styrene copolymers (ABS), polycarbonate (PC) and photopolymer by 3D printing, we obtain the force displacement curve of the three by the experimental instrument, as shown in the Fig. 4.

According to the Fig. 4, we get the maximum bending force of the hand orthosis of the photopolymer is 317.4 N, the PC is 143 N and the ABS is 95.4 N.

According to the reference [2], patients' maximal muscle tone is about 49 N, all the maximum bending force is greater than the patient's muscle tone, so these hand orthosis can be applied to the rehabilitation treatment. However, combining with material prices and other characteristics, for example the PC requires FDM printer's head temperature must be 230 °C, we choose ABS material.

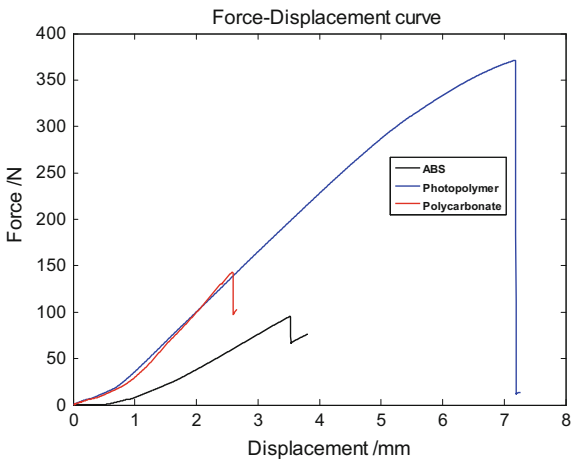


Fig. 4 Result of the mechanical property test

4.2 Construction of Finite Element Model

The hand orthosis is customized by the human hand structure, its curved surface structure and shape of cross section can not be simplified by the general material mechanics method, so use the finite element analysis software (Ansys) to calculate the bending strength of the middle finger of the ABS material.

Established finite element analysis model through the following steps.

(1) Apply finite element modeler and static structure modules in the project, then connect them for sharing model.
(2) Import the hand orthosis model of STL format into the finite element modeler, then click initial geometry and select small triangle faces as a components. Based on the rehabilitation therapy and contact forms between hand and the hand orthosis, create three components. The location of components are fingertip, finger root and the first interphalangeal joint.
(3) Set material parameters of the hand orthosis in the engineering data, including the density of 1020 kg/m^3, the Poisson's ratio of 0.394 and the Young's modulus of 200 MPa.
(4) Add one force and two fixed support in the static structure and the geometry of each one is these components as shown in the Fig. 5.

4.3 Result of Finite Element Analysis

Due to the complex cross section of the middle finger, the bending strength calculation is complicated, so we use Anasys to calculate. On the basis of the mechanical property test of the hand orthosis, we get the force when the middle finger broken, which is the maximum bending force, then get the bending strength by the calculation with this force.

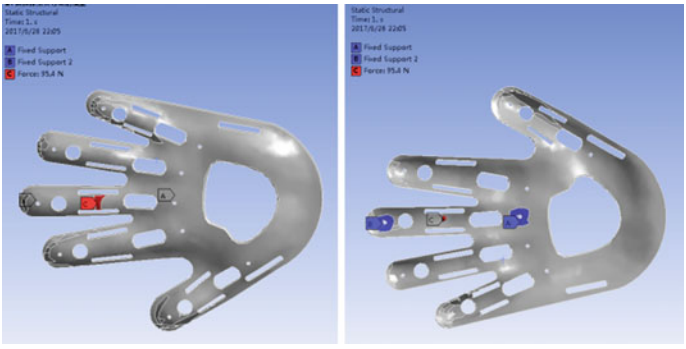


Fig. 5 Three components' location

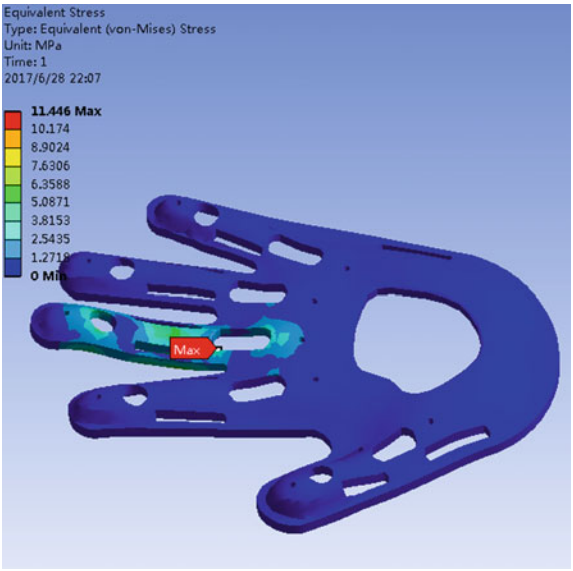


Fig. 6 Result of the finite element analysis

Thus set the force in the finite element analysis model to 95.4 N, after calculating, the bending strength is 11 MPa as shown in the Fig. 6. Much less than the allowable stress of the ABS material 24.5 MPa, so the hand orthosis is safe and reliable.

5 Conclusion

In order to achieve precision treatment of stroke patients, this paper used 3D printing technologies to make hand orthosis based on the size of the patient's own hand, then according to its clinical rehabilitation process, built a simplified model of middle finger, then made analysis of mechanics experiment and finite element analysis. The results show that the hand orthosis of the ABS material can be applied to the clinical rehabilitation treatment since it is not only made easily but also reliable and affordable.

The method and the finite element model proposed in this paper can be used for its future lightweight design, such as thickness optimization or the position and size of its holes.

References

1. Tang CZ, Jia J (2013) Research on the application of functional obstacle rehabilitation AIDS in stroke. China J Rehabil 28(4):252–254
2. Bao J, Chen J, Chen XL (2012) The influence of the combination of rehabilitation therapy on finger spasm after stroke. J Front Med (32):96–97
3. Scherling E, Johnson H (1989) A tone-reducing wrist–hand orthosis. Am J Occup Ther Official Publ Am Occup Ther Assoc 43(9):609–611
4. Gabriele B, Sami H, Alberto S (2016) A critical analysis of a hand orthosis reverse engineering and 3D printing process. Appl Bion Biomech 2016:8347478
5. Thibaut A, Chatelle C, Ziegler E et al (2013) Spasticity after stroke: physiology, assessment and treatment. Brain Inj 27(10):1093
6. Xu D, Jia J, Yang Y, et al. (2011) A report of 1 case of cerebral apoplexy after treatment with a combination of rehabilitation and rehabilitation therapy. Chin J Rehabil Med 26(11): 1065–1068

Fault Classification and Degradation Assessment Based on Wavelet Packet Decomposition for Rotary Machinery

Zhe Li, Viggo Gabriel Borg Pedersen, Kesheng Wang and Yafei He

Abstract This paper presents a novel method for fault classification and degradation assessment in rotary machinery through wavelet packet decomposition and data-driven regression methods. Wavelet Packet Decomposition is applied to extract the coefficient and energy based features from vibration signals. During the experiment, we used several machine-learning methods, including Artificial Neural Networks, Support Vector Machine, and K-Nearest Neighbor Classification for degradation assessment and compared the numerical results.

Keywords Fault classification · Degradation assessment · Wavelet packet decomposition · Rotary machinery

1 Introduction

As the key equipment in many production fields, rotating machinery covers a very broad range of industrial equipment and plays a momentous role in manufacturing application. It is one of the most common classes of mechanical equipment and generally may operate under a tough working environment, which make it vulnerable to types of faults. Moreover, these faults may cause equipment to break down or degrade certain machinery performance like geriatric location, manufacturing quality and operation safety [1].

In most cases, the subsystem in rotary machineries like gear transmission systems are not accessible or hard to inspect the failures directly due to restrictions of disassembly, huge machine size or environmental limitations [2]. Therefore, how to achieve fault detection, classification and degradation assessment failures in rotary machinery is always a hot issue in the research field of mechanical maintenance.

Z. Li (✉) · V. G. B. Pedersen · K. Wang
Norwegian University of Science and Technology, NO 7491 Trondheim, Norway
e-mail: zhe.li@ntnu.no

Y. He
Shanghai Polytechnic University, Jinhai Road, Pudong, Shanghai 2360, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_54

Many intelligent approaches for diagnosis or prognosis in rotary machine have been proposed and researched in the recent years [3–5]. Lin et al. [3] applied crossover characteristics to extract failure features from nonlinear data to detect faults for rotary machine. Wang et al. [5] proposed a method to selective ensemble neural networks for faults classification in rotary machine. Lu et al. [4] introduced a stacked denoising autoencoder to estimate the health condition of rotary machinery components. All these proposed methods have contributed greatly and achieve certain targets in relevant experiment. However, it still lacks approaches which can accurately classify the failures and assess degradation of certain components or performance for rotary machinery. Therefore, this paper proposed a novel approach based on Wavelet Packet Decomposition (WPD), Artificial Neural Networks (ANN), and K-Nearest Neighbor Classification (KNNC) for fault classification and degradation assessment in rotary machinery.

The rest of this paper is organized as follows: Sect. 2 briefly discusses the importance of vibration condition monitoring in rotary machinery. In Sect. 3, WPD is introduced to extract features from vibration signals for fault diagnosis and prognosis. Sections 4 and 5 describe the set-up and numerical result of the experiment. A comparison is also made with the proposed methods. Conclusions are summarized in the last section of this paper.

2 Vibration Condition Monitoring

The art of anticipating failure in rotary machinery by means of monitoring vibration is widely used in industry as almost 80% of common rotating equipment problems relates to misalignment and imbalance detectable by vibration monitoring. The measured vibration levels will change when a rotary machine has a defect. Vibrations caused by the defects occur at specific vibration frequencies, characteristic of the components, their operation, assembly and wear. The measured vibration levels indicate the severity of the defects [6].

Evaluation criteria for machine vibration is dependent upon a wide range of factors and the criteria adopted vary significantly for different types of machine. According to ISO standard 20816, there are three primary vibration quantities: displacement, velocity and acceleration [7]. However, in most situation, it is hard to give absolute vibration tolerances for any given machine. There is thus an obvious risk of judging measured vibration levels, dangerous when they are not, or the opposite—not dangerous when they are dangerous. Human experience and interpretation of the measured values still plays an important role in vibration condition monitoring. Therefore, how to apply machine-learning approaches for decision making with higher accuracy in vibration condition is always a hot issue.

Electronically measured raw signals are transformed in such a way that levels of these quantities describe the condition of a given rotary machine. Raw vibration

signals are transformed using analyzing techniques such as Fast Fourier Transform (FFT), Short Time Fourier Transform (STFT), Wavelet Packet Decomposition (WPD), empirical mode decomposition (EMD), Hilbert-Huang transform (HHT) etc. A common dilemma when analyzing vibration data from rotary machinery is to determine the vibration level acceptance criteria. This paper proposed a novel method based on WPD to analyze vibration data for fault classification and degradation assessment in rotary machinery.

3 Wavelet Packet Decomposition

WPD is a very useful tool to analyze vibration signals. In numerical analysis, the essence of WPD is a wavelet transform where the discrete-time signal is parsed through more filters than the discrete wavelet transform, which can provide a multi-level time-frequency decomposition of signals [8]. It is extended from the wavelet decomposition (WD) and includes multiple bases and different basis, which can result in different classification performance and cover the shortage of fixed time–frequency decomposition in Discrete Wavelet Transform (DWT) [9].

In DWT, the original signal will first pass through two complementary filters and emerges as approximation coefficients and detail coefficients, which includes the low frequency and high frequency information about the original signal respectively. The approximation coefficient will further split into a second-level approximation coefficients and detail coefficients. This process may repeat according to the number of decomposition layers. Figure 1 shows the 3-layer structure of signal based on DWT, where approximation coefficients and detail coefficients are labeled as A and D respectively.

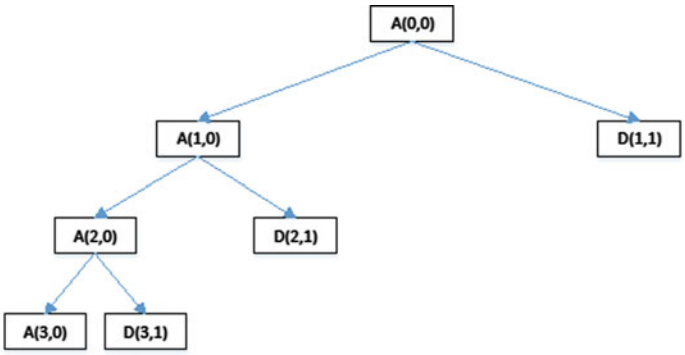


Fig. 1 3-layer structure of DWT

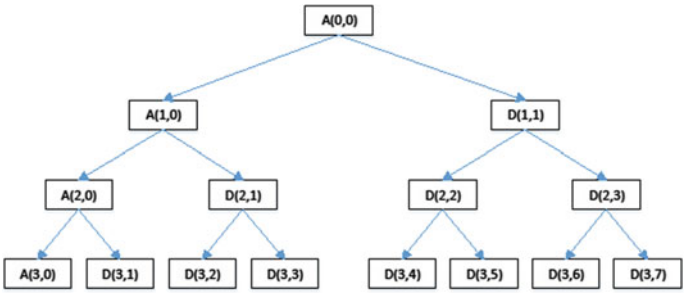


Fig. 2 Wavelet packet decomposition

However, WPD decomposes the detail and approximation coefficients simultaneously. Therefore, WPT can construct a complete wavelet packet tree with the same frequency bandwidths in each resolution. WPD can lead to a complete wavelet packet tree as shown in Fig. 2. A wavelet packet is a function with three parameters, i, j and k, which are the modulation, scale and translation parameters respectively.

According to the space partition shown in Fig. 2, we label W_j^i to represent the i th subspace of wavelet packet at the j th scale, where $W_{j,k}^n(t) = 2^{-\frac{j}{2}}\omega^n(2^{-j}t - k)$, k is the shift factor and $k \in Z$ [10]. It satisfies with Eqs. (1) and (2).

$$\omega_{j,0}^n(t) = \sum_k h_0(k)\omega_{j-1,k}^i \qquad (n \text{ is odd}) \tag{1}$$

$$\omega_{j,0}^n(t) = \sum_k h_1(k)\omega_{j-1,k}^i \qquad (n \text{ is even}) \tag{2}$$

where $j, k \in Z, n = 0, 1, 2, \ldots, 2^i - 1, h_0(k), h_1(k)$ are low-pass and high-pass filters of wavelet packet. Then the original signal $f(t)$ can be represented according to j level wavelet packet decomposition as Eqs. (3) and (4). The wavelet packet component $f_j^i(t)$ can be obtained through a linear combination of wavelet packet function $W_{j,k}^n(t)$ and wavelet packet coefficients $c_{j,k}^i$.

$$f(t) = \sum_{i=1}^{2j} f_j^i(t) \tag{3}$$

$$f_j^i(t) = \sum_{-\infty}^{+\infty} c_{j,k}^i(t) W_{j,k}^i(t) \tag{4}$$

Different types of wavelet functions may cause different time-frequency structures, in this paper, Daubechies 4(DB4) wavelet function has been chosen due to the good performance in estimations of the local properties of signals like

breakdown points [11], and the ability to derive a set of conventional and energy based features from signals [12]. During the experiment, we extracted the standard deviations of the coefficients and the percentage of energy corresponding to the approximation and details for fault classification and degradation assessment.

4 Set up and Data Collection

During the experiment, a Bently Nevada Rotor Kit is used to simulate the real working condition of rotary machinery. The sleeve bearing is equipped with 3 accelerometers of Kistler 8702B100, mounted in X, Y, Z direction, to measure the vibration signals from the test rig, as shown in Fig. 3. The sampling frequency is 4096 Hz and the maximum revolving speed is 4000 rpm. The bearing block is tightened down to foundation and can be loosened during the experiment. Rub generator and mass adjustable load can be modulated to simulate types of failures. The vibration monitoring refers to a zero position of the test rig. In this position, signals from the accelerometers are recorded and stored.

During the experiment, we injected three types of failures, bearing loosen, main spindle friction, and load imbalance. For each type of failure, the vibration signals will be measured at different failure degradation and rotating speed by means of acceleration meters and proximity sensors. Figure 4 shows the raw vibration signals and the wavelet coefficients and energy features.

After extracting the wavelet coefficients and energy features from vibration signals, Back-Propagation Neural Networks and Support Vector Machine Regression were applied as data-driven models to classify the failures and estimate the degradation.

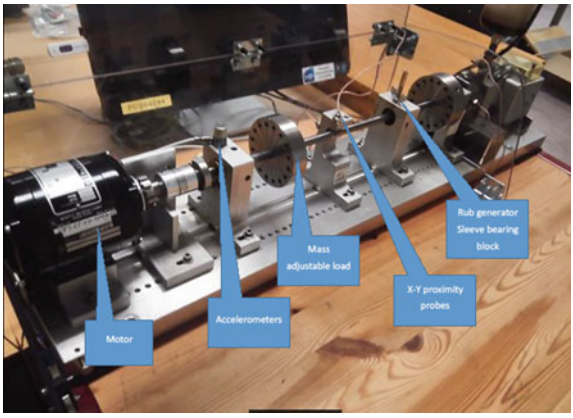


Fig. 3 Bently Nevada rotor kit

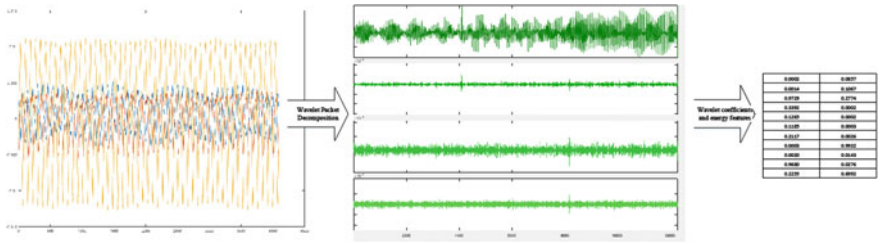


Fig. 4 Vibration signals and the wavelet coefficients and energy features

5 Numerical Result

During the experiment, a selection of 10,169 samples were applied to establish the data-driven models for failure classification and degradation assessment, respectively. In the failure classification stage, 6033 samples of data are used to train the neural network, and trained network is tested through other 784 samples. All the training samples and testing samples are selected randomly. Figure 5 shows the classification result from the Back-Propagation Neural Network (BPNN) through the wavelet coefficients and energy features. According to the test result, only one sample in the 784 testing samples is misjudged, which means the correctly classification rate is about 98.7%. Since the result from BPNN is already satisfied, we did not use other models to compare in this stage.

During degradation assessment, BPNN, Support Vector Machine Regression (SVMR), and KNNC are applied to estimate the degradation respectively. In this stage, we selected 3089 samples as training samples to train the BPNN and SVMR, and 310 samples as testing samples respectively. All these samples are collected under normal working condition or the failure of load imbalance with different degrees of degradation. The degree of degradation is adjusted by adding weight on the mass adjustable load from 0 to 800 g. Therefore, the inputs of the data driven

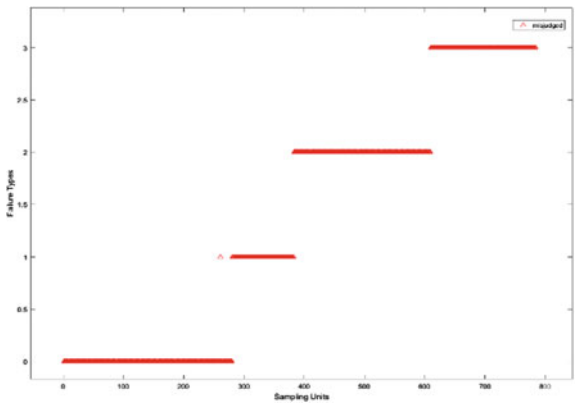


Fig. 5 Classification result from the back-propagation neural network

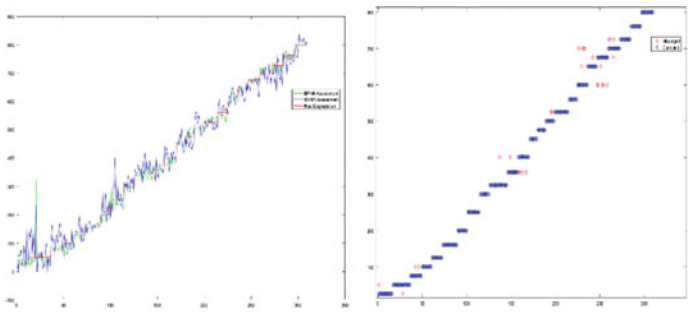


Fig. 6 Result of BPNN, SVMR, and KNNC

model are the coefficient and energy based features extracted from WPD, and outputs are the estimated weight on mass adjustable load, which may cause load imbalance. Figure 6 show the testing result of BPNN, SVMR with Gaussian Function and KNNC.

According to the testing result, the Mean Square Error (MSE) during assessment of BPNN and SVMR is about 424.28 and 1275.18 g, respectively. KNNC shares the best performance with 306.05 g in MSE.

6 Conclusion

This paper presents a new method for fault classification and degradation assessment in rotary machinery. At first, Wavelet Packet Decomposition is used to extract the coefficient and energy based features, which can represent the current condition of the rotary machinery. Then several data-driven models are used during the experiment to test their ability to detect, classify and estimate the degree of failures. According to the numerical results, as one of the most popular data-driven method, BPNN has the ability to deal with fault classification with the help of WPD for rotary machinery. However, as to the degradation assessment, we applied several data-driven methods, including BPNN, SVMR, and KNNC. KNNC shares the best performance in the experiment.

References

1. Lei Y, Lin J, He Z, Zuo MJ (2013) A review on empirical mode decomposition in fault diagnosis of rotating machinery. Mech Syst Signal Process 35(1):108–126
2. Yang Y, Dong X, Peng Z, Zhang W, Meng G (2015) Vibration signal analysis using parameterized time–frequency method for features extraction of varying-speed rotary machinery. J Sound Vib 335:350–366

3. Lin J, Chen Q (2014) A novel method for feature extraction using crossover characteristics of nonlinear data and its application to fault diagnosis of rotary machinery. Mech Syst Signal Process 48(1):174–187
4. Lu C, Wang Z-Y, Qin W-L, Ma J (2017) Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. Sig Process 130: 377–388
5. Wang Z-Y, Lu C, Zhou B (2017) Fault diagnosis for rotary machinery with selective ensemble neural networks. Mech Syst Signal Process
6. Scheffer C, Girdhar P (2004) Practical machinery vibration analysis and predictive maintenance, Elsevier
7. ISO 20816-1 (2016) Mechanical vibration—Measurement and evaluation of machine vibration
8. Zhang Y, Liu B, Ji X, Huang D (2016) Classification of EEG signals based on autoregressive model and wavelet packet decomposition. Neural Process Lett 1–14
9. Xue J-Z, Zhang H, Zheng C-X, Yan X-G (2003) Wavelet packet transform for feature extraction of EEG during mental tasks. In: Machine learning and cybernetics, 2003 international conference on, IEEE, pp 360–363
10. Ting W, Guo-zheng Y, Bang-hua Y, Hong S (2008) EEG feature extraction based on wavelet packet decomposition for brain computer interface. Measurement 41(6):618–625
11. Ferreira CBR, DbL Borges (2003) Analysis of mammogram classification using a wavelet transform decomposition. Pattern Recogn Lett 24(7):973–982
12. Murugappan M, Ramachandran N, Sazali Y (2010) Classification of human emotion from EEG using discrete wavelet transform. J Biomed Sci Eng 3(04):390

Research on Assembly and Disassembly of Reducer with the Combination of Virtual and the Actual Reality

Wenhua Zhu, Bao Cai, Kunju Shi and Sai Liu

Abstract The traditional practice teaching mostly adopts the mode of teacher's on-the-spot operation; however with the development of information technology, virtual simulation technology has been gradually applied in practice teaching. Shanghai Second Polytechnic University aims to train students' engineering quality and ability of comprehensive application of technology, using virtual simulation technology, in accordance with the general idea of "combination of virtual and reality", exploring a new model of practice teaching. Taking cylindrical gear reducer as object, applying virtual reality and augmented reality technology, developing new practice teaching resources, the combination of virtual and reality reducer is helpful for students to improve their practical ability, which can train students' professional consciousness and engineering quality.

Keywords Practice teaching · Virtual simulation · Augmented reality
Virtual reality combination · Reducer

1 Introduction

Reducer is general mechanical equipment, the purpose of using it is to reduce the speed and increase the torque, and its application areas involve construction machinery, metallurgy, shipbuilding, and power, petrochemical and other industries. Many universities have targeted courses on the principles, structure and performance of teaching. The traditional teaching method uses the reducer curriculum design method or adopts the reducer model engineering training method; however, it is often limited by the limited time for students to practice and the shortage of educational resources. In addition, with the rapid development of

W. Zhu (✉) · B. Cai · K. Shi · S. Liu
Engineering Training Center, Shanghai Polytechnic University, Shanghai, China
e-mail: whzhu@sspu.edu.cn

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_55

computer technology and information technology, the adoption of more innovative teaching methods has become an important means to overcome the technical difficulties of experimental teaching. The emergence of Virtual Reality can effectively solve many problems in traditional teaching.

The development of virtual simulation technology has been applied more and more, this paper summarizes the construction work of Chinese national virtual simulation experimental teaching centre [1, 2] in recent years, many Chinese universities are building a national virtual simulation laboratory Yun et al. [3] discussed the application of virtual simulation of large scale equipment. In 2005, Yu et al. [4] developed a virtual training system for CNC lathes, Zheng et al. [5] Applied virtual reality to civil engineering and shelter. After 2010, the application research based on VR was normalized, Yucheng [6] used VR technology to repair and display ancient ceramics, Quanwei [7] simulated the design of a pumping unit by virtual reality, Feng [8] makes reasonable suggestions on the application of VR technology to the elderly and the handicapped, Zhang [9] simulated the urban ground rail transit project, and explored the application of virtual reality technology in urban traffic, Li [10] developed a space station cabin routing simulation system, Wang [11] developed a 3D virtual campus walkthrough simulation system.

Some achievements have also been achieved in the field of virtual simulation in education and teaching, Wang et al. [12] summarizes the status quo and development of virtual simulation experiment teaching of foreign universities, analyses the characteristics and trend of virtual simulation experiment teaching, Wang [13] puts forward feasible steps in applying virtual reality technology to higher vocational practical teaching, Fang et al. [14] developed a virtual reality teaching system based on coal mine field based on Virtools technology, Yang [15] explored the application of VR technology in the field of higher education, virtual simulation technology has broad research prospects in the field of education and teaching.

In this paper, taking disassembly and assembly of reducer as the object of study, combining virtual simulation technology with traditional engineering training course of reducer, the disassembly and assembly of reducer combined with virtual and reality situation is studied, z disassembly and assembly reducer system with interactive and immersive features is developed to improve the students' practical ability.

2 General Research Idea

Study on the traditional reducer and training courses, the course contents include the understanding of internal and external components of the reducer, the working principle of the entity as well as on the deceleration disassembly and assembly measuring device and so on, which requires the development of the curriculum system on the list frame, in order to form a general idea of the whole research, Fig. 1 is the overall framework of virtual assembly system of reducer. This paper first establishes a reducer 3D model for teaching as the material, and then the

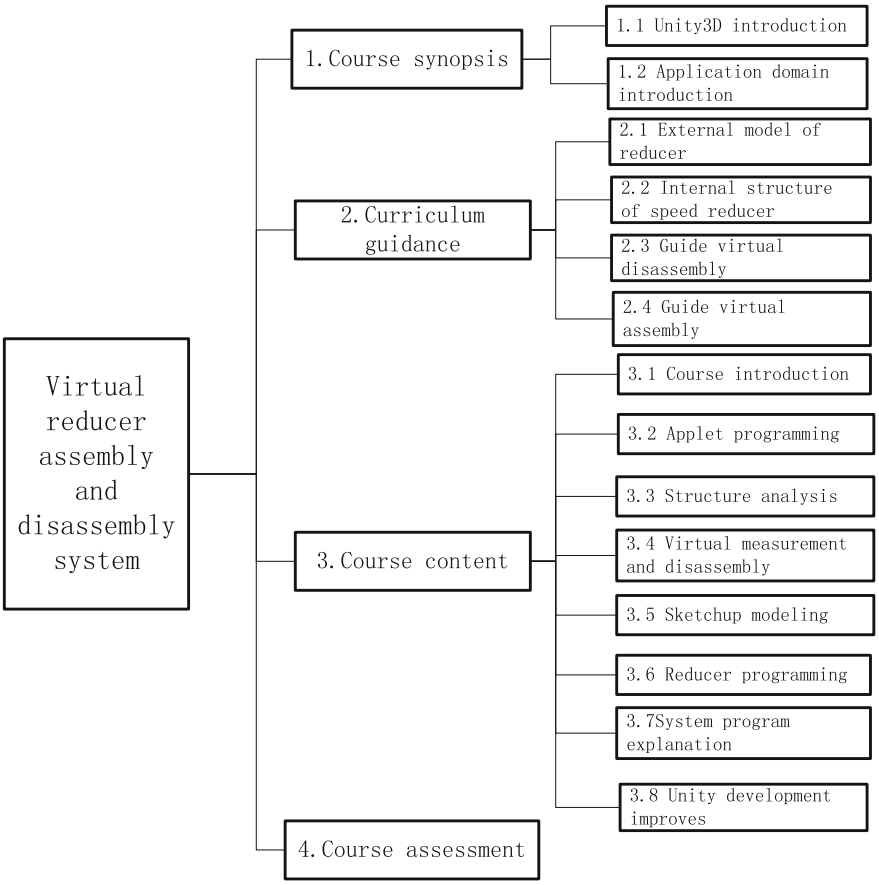


Fig. 1 Overall framework of the system

development of virtual reducer courseware system with Unity3D software, the use of virtual reality and augmented reality devices equipment Zspace and zView to develop the engineering training system combined with the actual situation of the reducer, its application in practical teaching.

3 Development of Courseware System Based on Unity3D

3.1 Preliminary Material Preparation

Unity3D software itself does not have a model, we must first set up according to the real corresponding three-dimensional model, Fig. 2 is a cylindrical gear for

Fig. 2 Reducer in kind

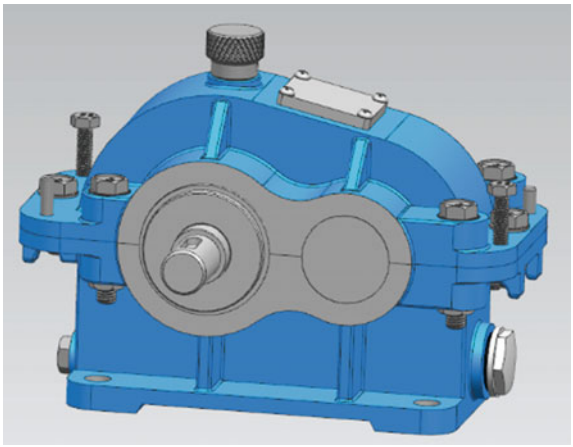


Fig. 3 3D model of reducer

students' practical engineering training of reducer, and Fig. 3 is on the reducer 3D model software under production proportion.

3.2 Development Process of Virtual Speed Reducer System

In Unity, 19 scenarios have been developed, which involve all the contents of the system framework, and call Unity3D internal functions for example:

1. The switch between scenes calls the Application function, and the statement is Application. LoadLevel ("Scene1").
2. Unity3D interactive system is one of the important characteristics of the interface in the button label can be interactive controls by calling the OnGUI ()

function, the corresponding method can call the function such as button position, font size and color, the design also curriculum objective questions test is also called toggle interactive control function.

3. The rotation and scale of the scene is realized by calling ray function and Lerp function. The function call statement is as follows:

   (a) ray = Camera.main.ScreenPointToRay(Input.mousePosition);
   (b) transform.rotation = Quaternion.Lerp (transform.rotation, mRotation, Time. deltaTime * Damping);

4. The movement and rotation of the model is implemented by Unity3D's free plug-in iTween animation library, which implements complex movements of sub models, such as:

iTween.MoveBy(gameObject,iTween.Hash
("y",5,"easeType","easeInOutExpo", "loopType", "none", "delay", 0.5));

Virtual reducer assembly and disassembly system part of the scene features are as follows:

General interface system (Fig. 4), including the composition of virtual assembly system of gear speed reducer: introduction to the course curriculum, instruction, and curriculum content and curriculum evaluation (in order to model parts easy to distinguish between the various parts of the development, rendering a different color).

Curriculum content interface (Fig. 5) is divided into eight courses, this part of the actual reducer with engineering training, while learning, so that students are more profound understanding of the whole learning process.

Virtual disassembly interface (Fig. 6), the reducer disassembly is divided into four modules: remove the upper body, remove the low speed shaft, high speed shaft disassembly and disassembly and disassembly process under the box body, can allow students to reset the scene repeated learning.

The curriculum test interface (Fig. 7) shows the knowledge points that appear in the system in the form of objective questions, and examines the results of the students' study.

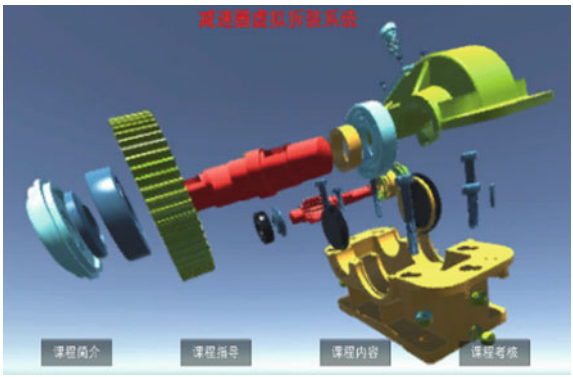


Fig. 4 General interface system

Fig. 5 Curriculum content interfaces

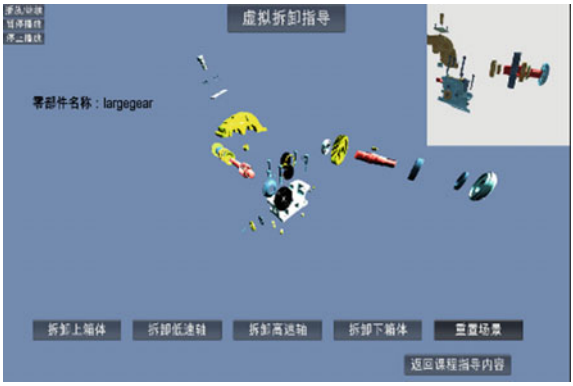


Fig. 6 Virtual disassembly interface

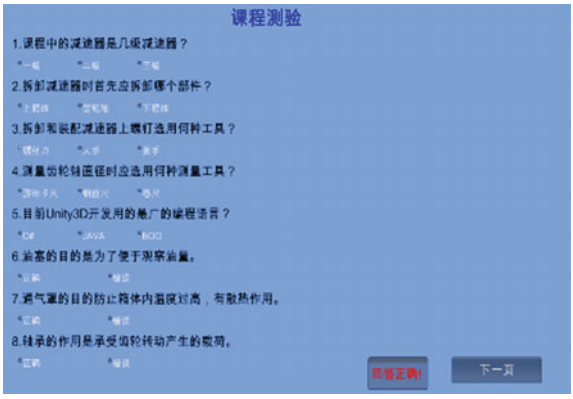


Fig. 7 Curriculum test interface

4 Development of Training System Based on Zspace

4.1 Zspace Profile

Virtual reality is a computer generated virtual environment that gives people a variety of sensory stimuli. Users should be able to interact with this environment in a natural manner, resulting in the illusion, immersion, immersive feeling of being in a corresponding real environment [16]. Virtual reality technology has the characteristics of multi sense, presence, interaction and autonomy. Zspace (Fig. 8) is the product of virtual reality technology.

Zspace system is a desktop virtual reality system which integrates the working environment of the real world, and realizes the free crossing of the reality and the virtual world. The technical core is the high fidelity stereoscopic display system, the low latency tracking system, and the software system. The core technologies of Zspace include stereo vision and direct interaction, as shown in Fig. 9.

Natural human–computer interaction—virtual interactive pen: Zspace interactive pen provides a natural way of human–computer interaction, and 6 degrees of freedom interactive pen provides sufficient protection to the experiencer.

Fig. 8 Zspace

Fig. 9 Zspace core technologies

Fig. 10 zView

Delayed position transmission—tracking glasses: Zspace tracking glasses use infrared technology to transfer experience location information to workstations with low latency.

In addition, the zView (Fig. 10), which is used with Zspace, has augmented reality capabilities that allow learners to share Zspace learning experiences easily with others.

4.1.1 System Development Characteristics

When using Zspace to develop virtual reality system, relative to Unity3D software, the most important thing is to reflect the interactivity, immersion and the combination of virtual reality and virtual reality.

4.1.2 System Interaction and Immersive Development

The curriculum development based on Unity3D interaction design is mentioned in the previous section by calling the OnGUI () function, which is realized through the click of a mouse button, although compared to the traditional multimedia courseware was intelligent, but to learn from both Chengdu and the degree of immersion can interact with personally on the scene, this section will try to call some function to achieve interactive objects of objects. When the reducer is close to the building

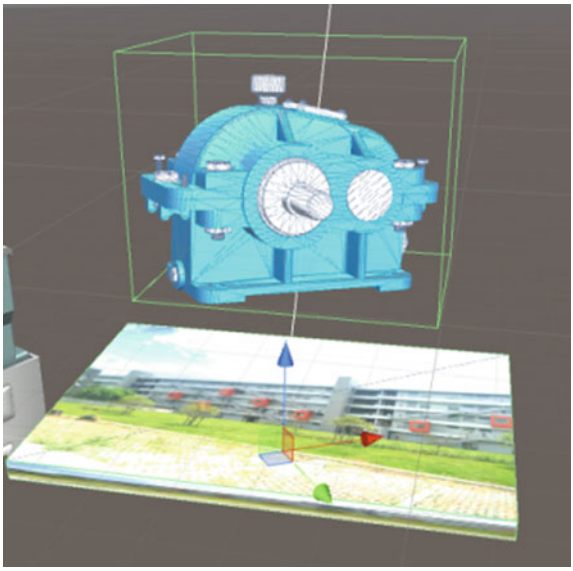


Fig. 11 Location between models

block, the distance between the models is calculated. If the distance between the two is less than the given value, the interactive function of the scene can be realized, and the key program for interaction is:

```
float dist1 = Vector3.Distance (reducer.position, Room.position);if (dist1 < 0.035)
{
Application.LoadLevel("interduction");
}
```

Combined with Fig. 11 and the upper segment program: firstly, the location information of ranging objects is obtained, then the distance function is called. If the distance between objects is less than 0.035, the interaction between scene scenes can be real.

4.1.3 The Display of Combination of Virtual and Reality

Virtual reality is the simulation of the real world, so when we develop virtual systems, we must conform to the logic of the real world in order to truly reflect the combination of virtuality and reality. In this paper, the virtual reducer disassembly system is developed. When the reducer is disassembled and developed, the disassembly process of the virtual reducer is developed according to the order of the actual reducer in Fig. 12. The disassembly process is shown in Fig. 13.

Figure 13 followed by removing the reducer upperbox, remove the lowspeed shaft and highspeed shaft and remove buttombox, if the student does not remove in order not to, for example in the disassembly box when the induction pen cannot

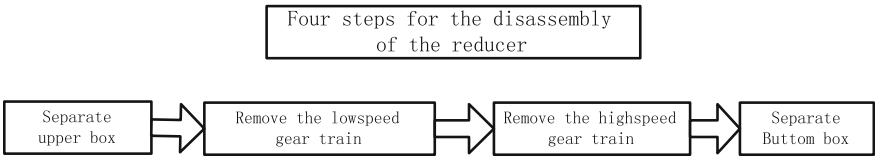


Fig. 12 Actual order of reducer disassembly

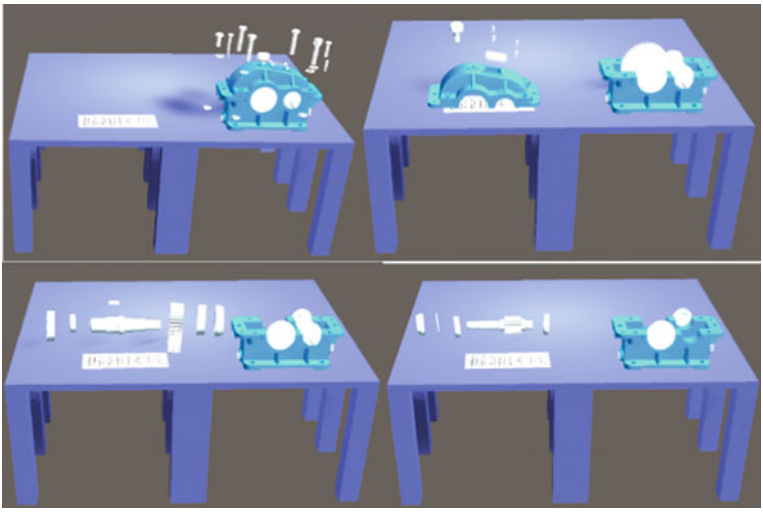


Fig. 13 Virtual reducer disassembly order

drag the low speed shaft, high speed shaft and a lower box model for disassembly, which is consistent with remove the ideological reality of reducer, will be in the virtual world and the real world of unified logic.

In addition, according to the actual reducer curriculum requirements, this paper has developed a complete virtual reducer disassembly system based on Zspace, including general interface, system interface, the structure of the reducer, reducer disassembly observation interface and reducer assembly interface, as shown in Fig. 14.

4.2 Augmented Reality Applications

Zspace as a three-dimensional display development equipment, in addition to the virtual reality function, but also has augmented reality function, zView devices that work with Zspace allow students to share the learning experience of virtual technology, Fig. 10 is a zView device for augmented reality, It is a high-definition

Fig. 14 Zspace development interface

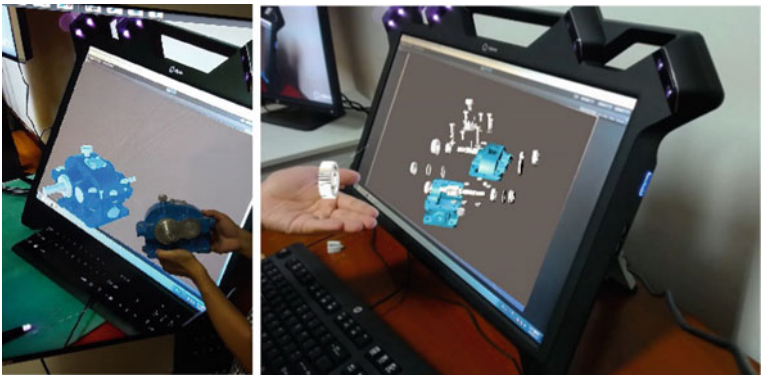


Fig. 15 Augmented reality

camera that overlaps the Zspace screen, which is designed to put the virtual world on the screen and interact with the real world. Figure 15 respectively deceleration augmented reality effects of external structure and internal structure, with the help of zView can make the model of reducer "out" screen to share learning experience and to realize the organic combination of virtual and reality.

5 Summary

1. The combination of reality and virtual reducer system is developed, used in the actual teaching process, compared to before the students' learning interest process has been greatly improved, the reducer structure and assembly of a more in-depth understanding of.
2. Using Zspace virtual reality equipment and zView augmented reality software, a virtual disassembly and disassembly system is developed, which greatly improves the interaction and immersion of virtual simulation, and the teaching effect is productive.
3. Virtual simulation technology and experimental teaching are integrated so that students can master the relevant content and skills better. Virtual interaction and virtual reality technology arouse the students' interest, stimulate students' desire for knowledge, and guide students to study independently, so as to achieve the purpose of applying what they have learned. The combination of virtual and actual reducer can help students improve their practical ability and cultivate their professional consciousness and engineering quality.

References

1. Xu J (2013) National virtual simulation experimental teaching center construction work summary and 2014 declaration proposed. Lab Res Explor 2014(08):1–5, 25
2. Sen W, Ping L (2016) Analysis of state level virtual simulation experimental teaching center in 2014. Lab Res Explor 04:82–86
3. Yun B, Sun J, Xiong X, Mao J (2015) Discussion on the role of virtual simulation experiment in open sharing of large scale instrument and equipment. Exp Technol Manag 10:119–121, 124
4. Yu Z (2005) Simulation training system of CNC lathe based on VR. Dalian University of Technology, Dalian
5. Zheng H (2008) The application and procedure of virtual reality technology in civil engineering and evacuation simulation realization of the fourteenth national engineering design of computer application conference. China Building Materials Industry Press, Hangzhou, pp 296–301
6. Yucheng C (2016) Research and application of restoration and demonstration of ancient ceramics base on VR technology. Sci Technol Commun 10:148–157
7. Quanwei C (2014) Research on the design and manufacture of pumping unit based on VR technology. Eng Value 19:37–39
8. Feng X (2016) Application of VR technology for the elderly and disabled people. Sci Technol 12:56
9. Zhang L (2014) Development and application of urban ground track transportation engineering simulation with VR technology. Sci Technol Perspect 31:214
10. Li M (2013) Design and research of cabin simulation system for space station based on VR technology. Harbin Institute of Technology, Harbin
11. Wang L (2016) Design and implementation of 3D virtual campus walkthrough simulation system based on Untiy3D. Chin Educ Inf 09:60–63

12. Wang W, Hu J, Liu H (2015) Present situation and development of virtual simulation experiment teaching in foreign universities. Lab Res Explor 05:214–219
13. Wang J (2011) Research on the application of virtual reality technology in higher vocational education. Chin Vocat Tech Educ 23:76–80
14. Fang N, Fan J, Cai D (2011) Development of mining virtual reality teaching system based on Maya and Virtools. Exp Technol Manag 11:128–130
15. Yang X (2014) Talking about VR technology and its application in higher education. J Changchun Normal Univ (Nat Sci Ed) 2:147–148
16. Zhu W (2007) Virtual reality technology and application. Intellectual Property Press, Shanghai Science Popularization Press, 10

Equipment Condition Monitoring System Based on LabVIEW

Zhongwei Rao, Bowen Feng, Lilan Liu and Yi Wang

Abstract In order to accurately predict the failure of operation equipment and make a reasonable decision timely. Adopting UDP communication protocol, using dynamic invocation and reentrant technology, the authors developed an equipment condition monitoring system based on Labview, and the feasibility of the system is verified by experiments. Experiment results show that the system can display and store the data fast and effectively and realize the time domain display and the spectrogram analysis. It's significant for the equipment fault diagnosis, accident prevention, and further improving production economic benefits.

Keywords Equipment condition · UDP protocol · Fault diagnosis Labview software

1 Introduction

With the continuous development of the digital construction of the workshop, the workshop management has become the prerequisite for improving the production efficiency and realizing the flexible, advanced and the safe production, the processing equipment is the foundation of the manufacturing line, and any tiny faults in the machine parts can reduce the production efficiency and result in safety hazard. Therefore, real-time monitoring and timely identification of equipment security risks are problems that need to be solved urgently.

Z. Rao · L. Liu (✉)
Shanghai Key Laboratory of Intelligent Manufacturing & Robotics, Shanghai University, Shanghai, China
e-mail: 18642260292@163.com

B. Feng
School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China

Y. Wang
School of Business and Management, Plymouth University, Plymouth, UK

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*, Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_56

Diagnostic analysis of equipment is usually vibration analysis, industrial vibration analysis technology is a detection tool to determine, predict and prevent the equipment failure, this technology involves in a variety of fields and variety of disciplines, such as sensor technology, computer technology, signal processing and analysis technology, accurate equipment vibration analysis will improve the reliability and efficiency of equipment, reduce downtime, eliminate electromechanical failure, through the front-end sensor information and programming, the system accesses data information to achieve multi-channel information parallel acquisition and complete the same time analysis of different devices [1], the collected signal outputs as voltage or current which is proportional to the vibration value in the form of "g" (gravity acceleration unit), the diagnostic platform achieves the most intuitive Waveform display and spectrum analysis, it's widely used in instrumentation control, data acquisition, data analysis and other fields [2]. This paper designs a set of vibration signal acquisition and analysis system for mechanical rotating equipment, the universal device is connected through the UDP communication protocol, and the original vibration signal is debugged and processed, and the data storage is completed to realize the transparent management of the production status of the workshop equipment.

2 Overall Scheme Design

As shown in Fig. 1, the complete vibration signal acquisition system is mainly composed of workshop equipment layer, data acquisition layer and real-time analysis layer.

In the workshop equipment layer, the system uses acceleration sensors. Through the portable magnetic seat to collect vibration signal for perceiving the workshop rotating equipment. In the data acquisition layer, in order to reduce the difficulty of wiring, using wireless communication technology to realize ad hoc network connection. Through the RJ45 interface and the UDP communication protocol, the synchronization acquisition of multi-channel bottom signal is realized. Through the fixed IP and port number, the data acquisition instrument is connected with the sensor point to point, realizing the data exchange between the host computer and the slave computer. In the final real-time analysis layer, the data center realizes the dynamic calling of server data through the programming of the computer software, real-time display of vibration waveform, real-time analysis of FFT/spectrum and storage of data are realized according to the device data sensed by sensors. Finally, human-computer interaction is achieved through friendly interface design.

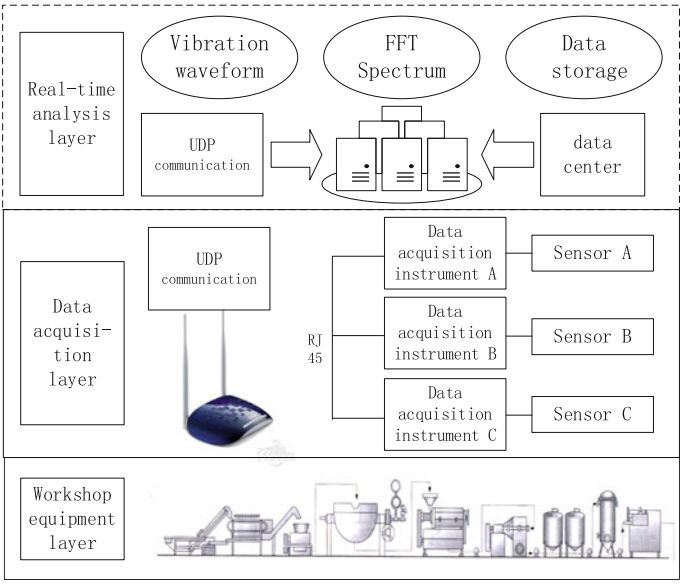


Fig. 1 System frame diagram

3 Key Technologies for System Design

The signal acquisition system is mainly composed of lower layer acquisition hardware, PC server and host computer software. The system requires real-time acquisition, high efficiency of analysis and low cost of testing, which put forward the very high request to the design of software and hardware.

3.1 Structure Design of Acquisition Instrument

Data acquisition device is the core of system hardware design, as shown in Fig. 2, the data acquisition instrument of system has two functions [3]: data acquisition and signal conditioning, and built-in power supply module in order to facilitate external sensor power supply. Completing the communication connection based on the Ethernet interface through the way of non-blocking to receive and transfer data.

As the workshop production environment is wretched [4], the vibration signal collected by acceleration sensors are often mixed with external interference. The data acquisition device will preprocess the mixed signal, through Voltage amplification filter module, Charge debug module, ICP debug module, 4–20 mA debug module and

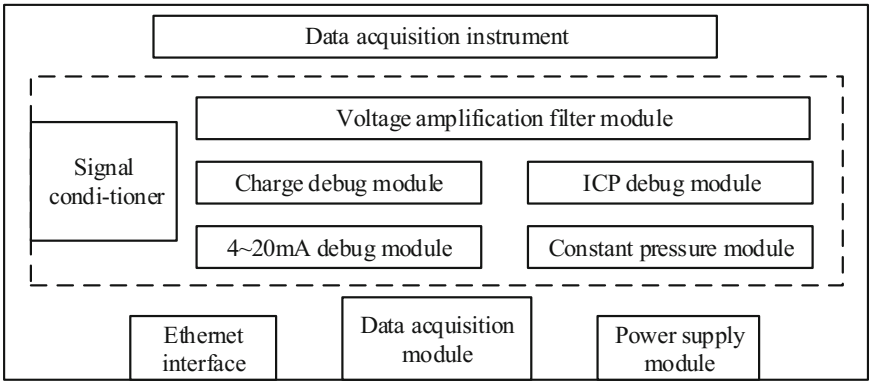


Fig. 2 Hardware structure diagram

Constant pressure module to achieve the signal amplification, isolation, filtering and multi-channel conversion, which greatly prevents signal interference and attenuation, and improves the reliability and integration of the data acquisition system.

3.2 Communication Design and Data Processing

Server-side software platform built-in rich data acquisition, storage, analysis and other toolkits, which support a variety of communication protocols [5], then, the equipment condition monitoring system through the dynamic calling technology, reentrant technology and database technology to complete multi-channel signal synchronization acquisition.

3.2.1 UDP and Multi-channel Synchronization Acquisition Technology

The bottom data acquisition device through the ready instruction upload content "Ready" and the handshake command to download "ok" to achieve handshake with the computer-side. The handshake signal is shown in Table 1, using the UDP protocol to complete the real-time vibration signal acquisition.

UDP as a common communication protocol provides a reliable network. Opening the UDP connection by the server name and port number. Data is written

Table 1 Handshake format

Serial number	Features	Transmission direction	Format/Content
1	Ready instruction	Upload	"Ready"
2	Handshake instructions	Down	"Ok"

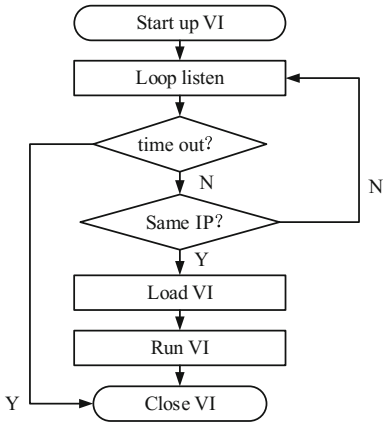


Fig. 3 The flow chart for dynamic invocation

to the server VI continuously. Through creating a Read UDP data VI to get the underlying packet. UDP connection is encapsulated into sub-VI, waiting for client connections. In the main VI, the flow of data is updated continuously through the matching of the listener IP and the timeout installed. Then, the data stream, in the specified network, is read and displayed by the read data VI. Continuous data acquisition is performed by additional cyclic structure function. Finally, turn off the network connection.

A major innovation of this system is the simultaneous acquisition of multi-channel signal, the core of synchronous acquisition is the implementation of the dynamic call of VI and the reentrant Technology. Dynamic call technology greatly saves system memory spaces, the sub-VI is automatically loaded and run only when the system needs it. The flow chart for dynamic invocation is shown in Fig. 3. In addition, the use of reentrant technology well avoids that an access violation occurs when multiple different threads call the VI at the same time, in the program. The core of the technology is to open a new data area on the VI that called and ensure that each reentrant VI has separate data space when called simultaneously by different threads [6], so as to achieve multi-threaded synchronous acquisition and improve the system efficiency.

3.2.2 Data Analysis and Storage

Acceleration sensors collect a great number of signal date so that analysis and storage choose fixed frequency date sample that is fixed length. The middle character of date package demonstrate clearly by using ASCII code that is hexadecimal. The formation of date package is "UPDATA + date package of hexadecimal + … + date package of hexadecimal + END". The shaking date is made up of high-byte and low-byte. Finally, it is transferred by decimal short integer character symbol. The storage of date is the important step of diagnosing equipment, which is the basis for post-operation workers to query.

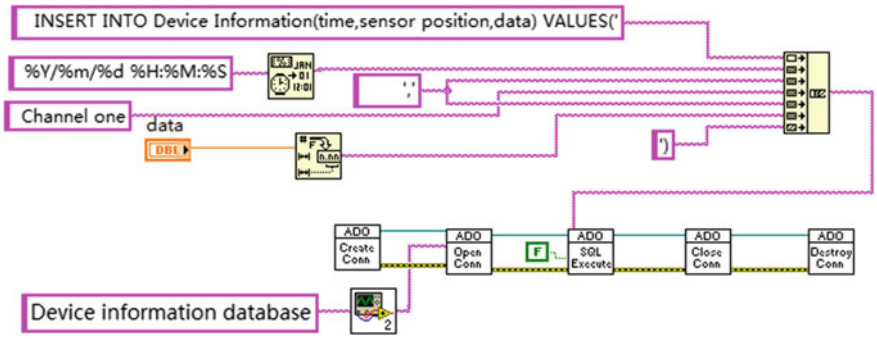


Fig. 4 The way date store

LabSQL toolkit, which is built in Labview platform itself, save date message to the sheet of Access database by the step of creating, opening, storing, closing and deleting so that the equipment can trace back easily. How the date store is shown in the Fig. 4.

3.3 Signal Analysis

The system mainly analyzes the operation of the workshop equipment from two aspects: time domain analysis and frequency domain analysis [7], to predict the equipment failure point and ensure the production safely and effectively.

3.3.1 Time Domain Analysis

Due to the complexity of the workshop situation, we must need to filter the vibration signal. LabVIEW's signal processing module provides a large number of filters [8], such as Elliptical Filter, Median Filter, Equi-Ripple High Pass and so on. The system uses Chebyshev low-pass filter for vibration signal to time domain analysis, the waveforms of the signal before filtering and after filtering are shown in the Figs. 5 and 6.

At the same time, the system use the filtered data to performs Mean, standard deviation and variance, the formula of variance is

$$D(X)=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}) \tag{1}$$

n represents the total number of data analysis points, x_i represents current value and represents data average.

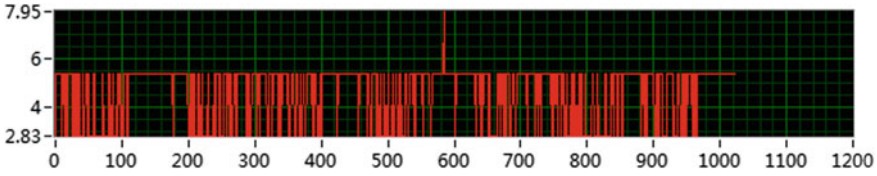


Fig. 5 The waveforms of the signal before filtering

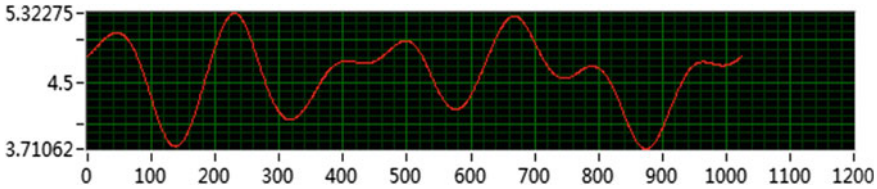


Fig. 6 The waveforms of the signal after filtering

The formula of standard deviation is

$$\sigma = \sqrt{D(X)} \tag{2}$$

Time domain analysis clearly reflects the relationship between amplitude and time, and helps determine the time of failure.

3.3.2 Frequency Domain Analysis

The frequency domain analysis of the vibration signal mainly reflects the alternative relation between the rotation frequency of the equipment and amplitude to predict the point of failure and make up the limitations of time domain analysis. The frequency domain analysis is mainly carried out the Fast Fourier Transform (FFT) of signal, so as to the time domain signal is converted into frequency domain signal, and finally output in the form of a bilateral spectrum conversion. The conversion format is shown in Fig. 7.

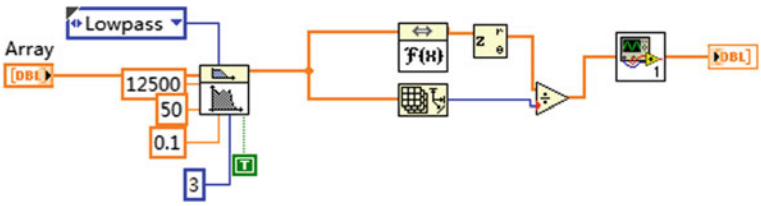


Fig. 7 The conversion format

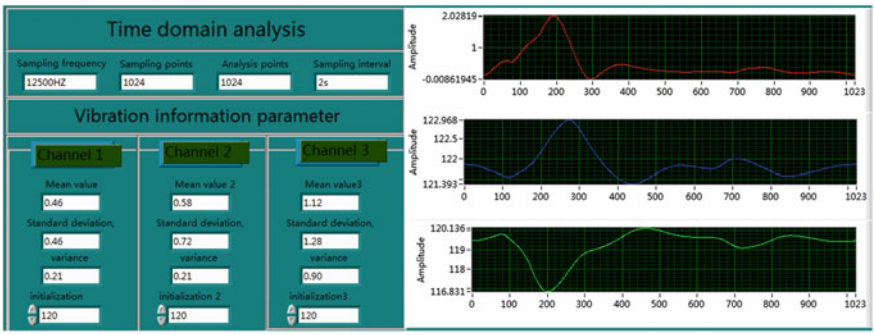


Fig. 8 System interface

The frequency domain analysis of the vibration signal well shows the position of each frequency component that from composite vibration signal, facilitating the development of reasonable countermeasures.

3.3.3 Example Validation

The system proposed in this paper has been successfully applied to the workshop of Shanghai Key Laboratory of Intelligent Manufacturing and Robotics, which solves the problem of data acquisition and analysis of multi-channel equipment.

The front-end sensor of the system is distributed in a number of sensing points of the EMCO Tum 242 CNC machine tool, we can know the operation of the equipment through the vibration signal acquisition and analysis, which solves the problems that the data interchanges inconveniently and the operation of equipment is not transparent, it greatly reduces the failure of production safety and increases the product processing accuracy. The system interface is shown in Fig. 8.

4 Conclusion

Based on LabVIEW, this paper designs a set of data sensing and analysis system for rotating equipment, through UDP protocol, dynamic call and reentrant technology, this paper realizes the simultaneous acquisition of multi-channel data and equipment condition monitoring. The main functions of this system include the time domain analysis and frequency domain analysis of signal, through the relationship of amplitude and time and frequency domain, we can predict the fault point and maintain it easily. The experimental results show that the accuracy and validity of the system are verified, the experimental results are in accordance with the actual situation of the equipment. And the host computer interface is simple and clear,

interactive convenience. The realization of the system provides a great reference value for the monitoring of production equipment vibration in the future.

Acknowledgements The authors would like to express appreciation to mentors in Shanghai University for their valuable comments and other helps. Thanks for the pillar program supported by Shanghai Science and Technology Committee of China. The program number is No. 17511104600.

References

1. Neuzil J, Kreibich O, Smid R (2014) A distributed fault detection system based on IWSN for machine condition monitoring. IEEE Trans Ind Inform 10(2):1118–1123
2. Rao Z, Yang S, Ma J, Wang S, Liu L (2017) Wisdom workshop information perception and analysis. Mod Manufact Eng 5:22–27
3. He Y (2010) Vibration signal acquisition system of engine based on LabVIEW. Manufact Autom 9:196–199
4. Li Y, Liao P (2015) Design of large rotor vibration signal acquisition system based on STM32. Instrum Tech Sens 12:65–68
5. Wu Yinmin, Han Y, Cui G-Q (2006) Analysis of vehicle vibration measurement base on virtual instrument. Chin J Sci Instrum 27(6):19–420
6. Li C, Jiao R-L, Chen J-T (2012) Parallel data acquisition system design based on LabVIEW. J Liaoning Tech Univ (Nat Sci) 31(1):89–91
7. Liu T, Hua S (2009) Vibrations signal analysis for inverter scroll compressor based on LabVIEW. Electron Measur Technol 32(7):116–118
8. Chen KR, Xiao TB, Chen JM (2012) The design of multi-channel synchronous data acquisition system of ceramic brick press based on LABVIEW. Hydraul Pneumatics Seals

The Algorithm Knowledge Base for Steel Production Process Optimization

Shuai Li, Lilan Liu, Yiping Wu and Yi Wang

Abstract The goal of algorithmic knowledge base is to solve the problems of various constraints and complicated process in the process of steel production. The optimal scheduling of resources, energy and technology is the core, and the optimization problem of production is solved quickly based on various intelligent optimization algorithms. Through the different optimization algorithms in the knowledge base to realize the optimization of various problems in steel production. Algorithms base collecting all kinds of optimization algorithm and application scenarios, establish intelligent algorithms base for iron and steel production process and providing intelligent decision support for intelligent steel production improving production efficiency and product quality.

Keywords Steel production process · Algorithm knowledge base · Data base

1 Introduction

The method base was developed on the basis of the base, which was developed in the 1970s with specialized computing software packages, called base [1]. The base is targeted, but there are limitations. The limitation is mainly manifested in the base by different user subroutine calls, each time to compile and link, modify the program base routines, to modify the user calling the program; in order to use the base, users must be familiar with the programming language and data rules. For the limitations of the base, there is a method base, method base system is a component

S. Li · L. Liu (✉)
Shanghai Key Laboratory of Intelligent Manufacturing and Robotics,
Shanghai University, Shanghai, China
e-mail: 997729619@qq.com

Y. Wu
Shanghai Baosight Software Co., Ltd., Shanghai, China

Y. Wang
School of Business and Management, Plymouth University, Plymouth, UK

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_57

of the expansion program, and they can be connected to a variety of database, and the application of the control system, the specific application requirements into the corresponding program system. In the middle of 80s, the method base system became more intelligent. In the middle of the 90s, the development of web-based DSS became an active area of research, and the method base system was heavily influenced [2]. At present, China's computer industry research and development of the method base system is further in-depth, which requires a long process, in its research process will always find new problems, find problems, solve the problem is a cycle of repeated process, in this process, people continue to improve and improve the method base system. Method base systems are a component part of decision support system, which includes method base, method base management system and method base dictionary. The method base system can be divided into three levels: base level, matching level and application level. The methods provided at the base level are called meta methods, which form the base set of methods base. The application level provides a mathematical model for end users with application problems. In the early computer literature, the form of the method base was called model base [3]. The matching level is between the basic level and the application level, and the basic method is synthesized into a special method, and the data is connected with the processing method. Method base is the tool of storage method module, which is composed of various common and flexible algorithms, which can be used to form various mathematical models. The method base management system is a software system of centralized control method module. The main functions include the nomenclature, classification, storage and invocation, method execution, data and method connection, and so on. The method base dictionary is information stored and the method itself (method category, function, scope, form and method of parameter call form), to complete the program login, to provide the necessary information to facilitate the learning process, connection method base and database etc. The purpose of this paper is to combine the optimization problem in the steel production process with the knowledge of the method base system, and to design the steel process optimization algorithm base system, which can be used to manage the steel production process algorithm uniformly and effectively. So that enterprises in the production process of different problems have a more convenient solution to improve production efficiency.

2 Demand Analysis

2.1 Analysis of Optimization Algorithms for Iron and Steel Production Process

The main processes of iron and steel production include iron making, steelmaking [4], continuous casting, rolling and so on. The energy consumption of iron and steel industry is the large energy consumption industry, because of the characteristics of

the steel industry itself, requires a high degree of consumption type of energy and resources, such as water, steam, electricity, coal and other resources, the number of Chinese annual energy consumption of iron and steel enterprises accounted for the industry energy consumption quantity is about 15%. In the energy consumption structure of China's iron and steel industry, coal accounts for 69.9%, electricity accounts for 26.40%, fuel oil accounts for 3.2%, and natural gas accounts for 0.5% [5]. In the process of iron and steel production, the energy consumption of iron and coal consumed by iron smelting is the highest. Hot rolling is the main part of steel rolling. The main processes include: heating, removal, edge grinding, rough rolling, hot rolled, finish rolling, laminar cooling and coiling. The main energy consumption is the consumption of the pre heating furnace and the energy consumption of the roughing mill and finishing mill in rolling process [6]. Therefore, the optimization of the process is mainly based on the control of reheating furnace and the optimization of slab parameters in the process of rolling. Secondly, before entering the steel rolling, the steel needs to pass through the converter continuous casting stage, and this part needs reasonable scheduling according to the order of the enterprise, so as to meet the demand of customers and the lowest energy consumption. In the steel plate forming, it is necessary to test the quality of the steel plate, so as to achieve the standard, so the quality inspection of the steel plate is also the focus of the production process. To sum up, in the process of steel production, we divide the application of the algorithm into the following items according to the optimization direction: optimization of process parameters, optimization of equipment parameters, scheduling optimization and steel plate quality inspection.

2.2 Functional Requirements Analysis

The complete optimization algorithm base system platform for users in the enterprise is convenience. So that users quickly learn the operation of the system. The functional requirements of the system mainly include the following aspects:

1. System Management
   The system management function mainly realizes to provide the user friendly interface, including user basic information management, password setting, authority setting, log management. The main user basic information to add, delete, save, query, modifies passwords, user permissions settings and the management and the log. Because the main function of the system management operation authority is relatively high, so the operating system management workers is mainly the system administrator, but the operation results can be presented to any participants, so the management system use case diagram as shown in Fig. 1.

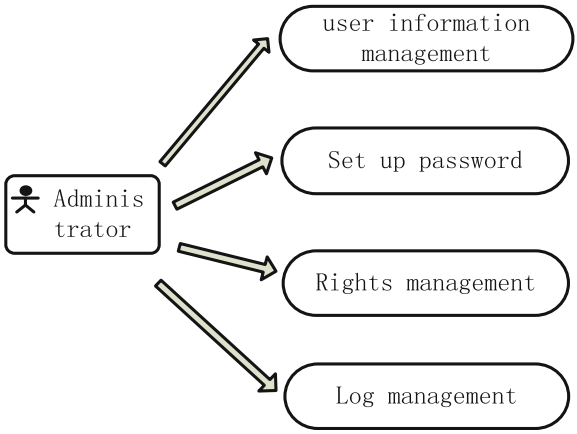


Fig. 1 System administration diagram

2. Data Maintenance
   Data maintenance mainly completes data initialization, data backup and data recovery operations, the relationship between specific use cases and the relationship between workers, participants and use cases, as shown in Fig. 2.
3. Data Dictionary
   The data dictionary mainly completes the main operation of the dictionary table in the system name, algorithm type and so on. In order to complete the algorithm name and the algorithm type, the concrete operation is shown in Fig. 3.
4. Algorithm Basic Information Management
   The main implementation of the algorithm, the basic information insert, delete, update, query, print, sort and other functions. The specific relationship is shown in Fig. 4.
5. Algorithm Code Management
   The main algorithm code insert, delete, modify, save, set, import, clear, print and other functions. A relationship between specific use cases is shown in Fig. 5.

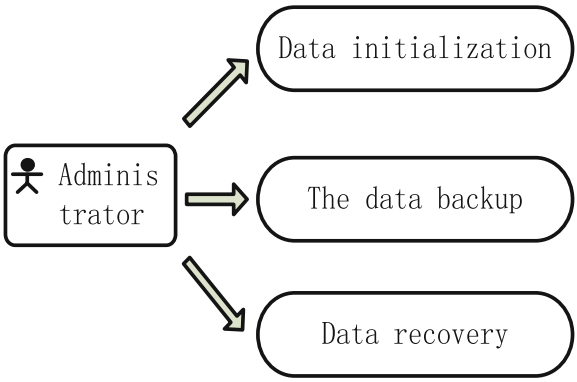


Fig. 2 Data maintenance case diagram

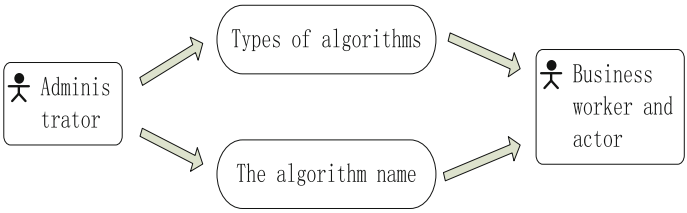


Fig. 3 Data dictionary diagram

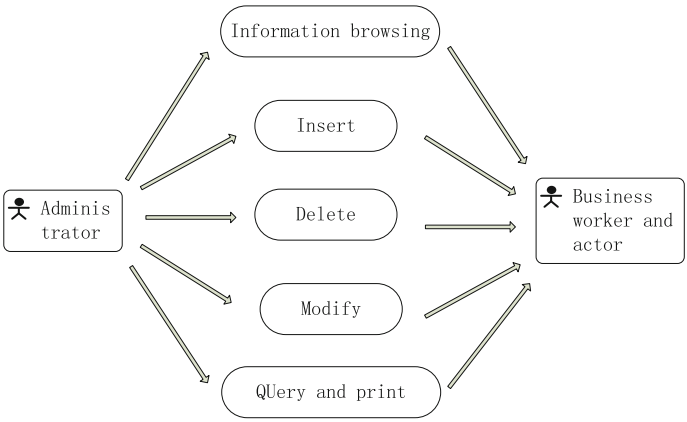


Fig. 4 The basic information of the algorithm

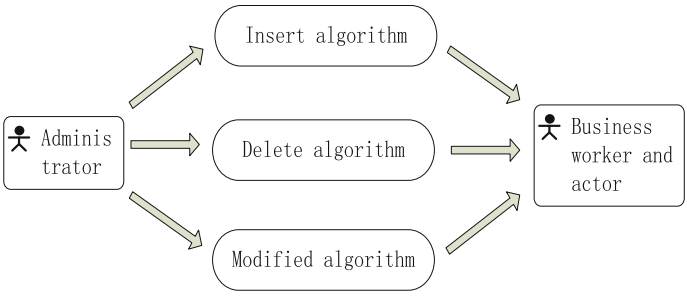


Fig. 5 Algorithm code management case diagram

3 Design of the Algorithm Base System

3.1 Overall Design of Algorithm Knowledge Base Structure

The maximum to meet the demand of the above analysis, due to the needs of different users are not the same, the system needs to meet the individual needs of each different, so based on the data dictionary should be complete, should be more to meet the property covers algorithm for demand; second is the maintenance of the system, the software will be divided into function different module to design and implement in the development process, which requires integration of the module can be very good, and need to maintain the interface between the modules and dependencies, to design in the software development process framework more flexible, scalability is better, otherwise it is difficult to maintain, and will increase the cost of maintenance. The quality of the system is also very important and needs great attention. Therefore, to do a good job of system analysis and design before the construction of the system will make the development more effective. The overall design is the most important part in the design process, through the demand analysis phase, the formation of optimization algorithms base system platform design idea, determines the subsequent development of all aspects of design, data structure algorithm base system, the overall design is normative, with a rational organizational structure.

The design of algorithm base for steel process optimization is based on the maximum convenience for users to manage and use the algorithm, so as to meet the needs of different problems in the process of iron and steel production. When the data dictionary model of the algorithm base is set up, the attributes of various aspects of the algorithm are collected, such as algorithm name, category, author, algorithm function description, and the function of compiling and executing the algorithm. Calling a different integrated development environment to run the algorithm, not only saves the system resources, but also allows users to quickly use the algorithm, user-friendly algorithm for horizontal and vertical comparison, and thus further familiar with the algorithm. The overall foreground background module design of the database is shown in Figs. 6 and 7.

The system management module includes the user (or user) basic information management, password settings, permissions settings, log management. Mainly realize the basic user information add, delete, save, query, password changes, user rights settings, and log management functions. The data maintenance module mainly includes data initialization, data backup in the data window to the database, the formation of spare parts files and data recovery from the backup files. Data dictionary module includes a list of commonly used data: such as the name of the algorithm, algorithm categories, user rights and other dictionary tables, this module is mainly for the convenience of users and set up. The basic information management module of the algorithm mainly realizes the function of the basic information of the algorithm, such as inserting, deleting, updating, querying, printing and sorting. While deleting the basic information of the algorithm, the code files of

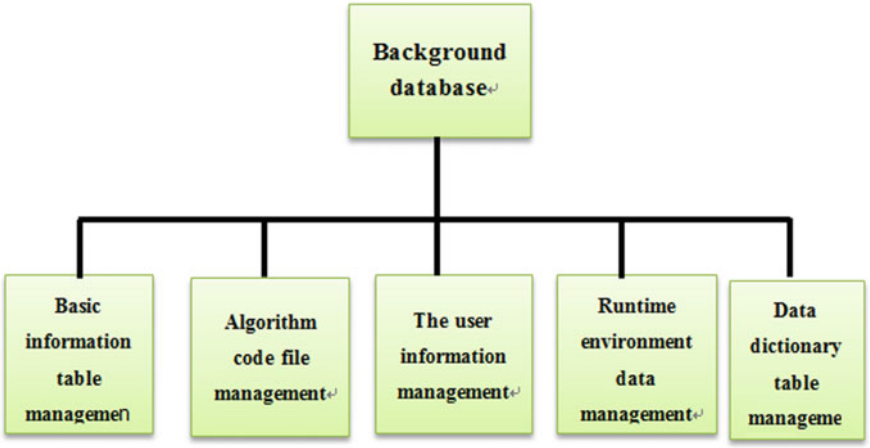


Fig. 6 Overall design background management system function module

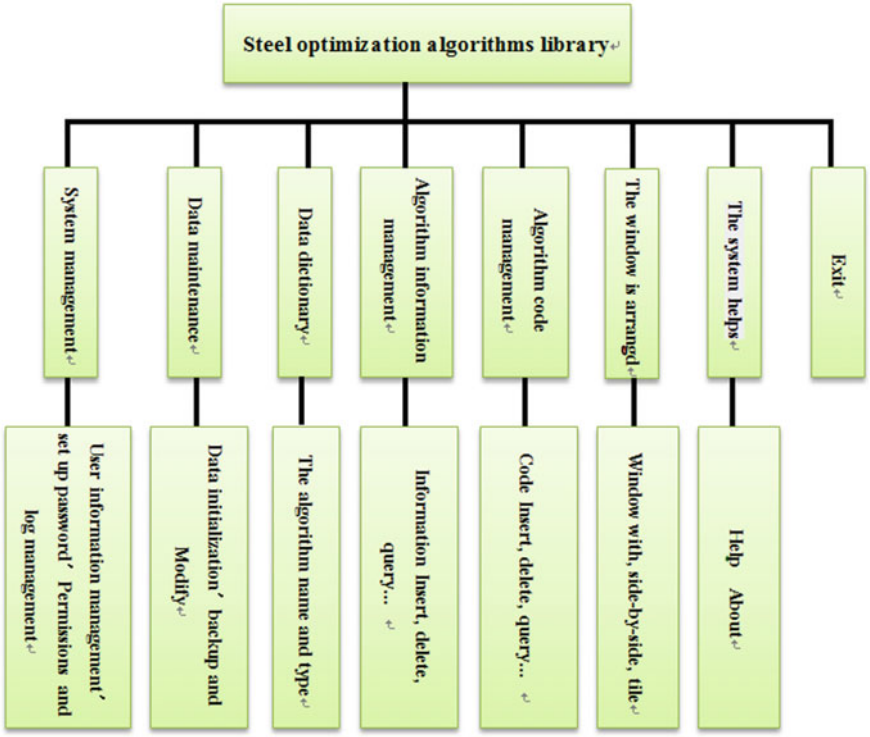


Fig. 7 Front desk management system module design drawing

the algorithm can also be deleted from the background. The code management module is one of the key modules of the system, mainly to achieve the insertion algorithm code (i.e. adding new algorithm code), algorithm code to delete the data window and the corresponding file, modify the existing algorithm code, save algorithm code, set (will be displayed in the data window code algorithm name set to the selected state), import (algorithm code file into the data window), clear (the algorithm code database contents all clear, so this base includes only the structure, no record), print (to copy the data window in the algorithm code) and other functions. Window management module are arranged on the window, due to open multiple windows during program execution, in order to make the interface looks beautiful and orderly, can choose the window arrangement. Such as: parallel, horizontal, flat, cascade etc. The help module provides a detailed help function for users to use, and users can see the help directly by helping them understand how the system is used. The exit module completes the function of exiting the whole system. The system is written in visual studio environment using c#, and the background database is implemented by SQL server 2005.

3.2 Database Design

1. Database Selection
   We usually refer to database systems as database engines for application systems, because they are the core of the entire application system. In the design and implementation of this system, SQL Server is chosen as the database of this system.SQL (Structured Query Language) [7], which means structured query language, is a structured query language between relational algebra and relational calculus. Its main function is to establish contact with various databases and communicate with each other. ANSI (American National Standards Institute) stipulates that SQL is the standard language of relational database management system. SQL statements can be used to perform a wide variety of operations, such as updating data in a database, extracting data from a database, and so on. SQL language data query, data manipulation, data definition and data control functions, according to the need to gradually change mode, and it does not affect the operation of the database, so that the system has good extensibility [8]. At the same time, the method of set operation is not only the object of operation, the result of finding can be the set of records, but also the object of inserting, deleting and updating operation at one time. So, finally choose SQL Server 2005 to achieve the background database design.
2. Database Structure Design
   The design of this database mainly contains 4 tables, which are user information table, administrator information table, algorithm basic information table and algorithm code import table, as shown in the following Tables 1, 2, 3 and 4.

Table 1 User information

Field name	Data type	Length	Explanation
Username	Char	10	User name
Password	Char	10	User password
Grade	Int	4	Grade

Table 2 Administrator information

Field name	Data type	Length	Explanation
Admin	Nvarchar	50	Administrator name
Password	Nvarchar	50	Administrator password

Table 3 Algorithm basic information

Field name	Data type	Length	Explanation
Algorithname	Varchar	30	Algorithm name
Ck	Bit	1	Algorithm is selected or not
Kind	Char	10	Algorithm kind
Author	Char	10	Author
Filename	Varchar	30	File name
Filecodepath	Varchar	60	Code path
Runpath	Varchar	80	Code file run path
Description	varchar	100	Description of algorithm

Table 4 Algorithm code

Field name	Data type	Length	Explanation
bh	Int	4	Line number of code
Code	Varchar	100	Content of each line statement

4 Application of Knowledge Base

4.1 Development Environment of Knowledge Base

Hardware development environment conclude: CPU: Pentium4-2.0G, memory256 M, Hard Disk40G, displayer 17", resolution 1024*768, refresh rate 85 Hz, network card1 00 M. Software development environment conclude: Operating System windows 7,32 bit, database SQL server 2005, development platform Microsoft Visual Studio 2015 and the language is c#.

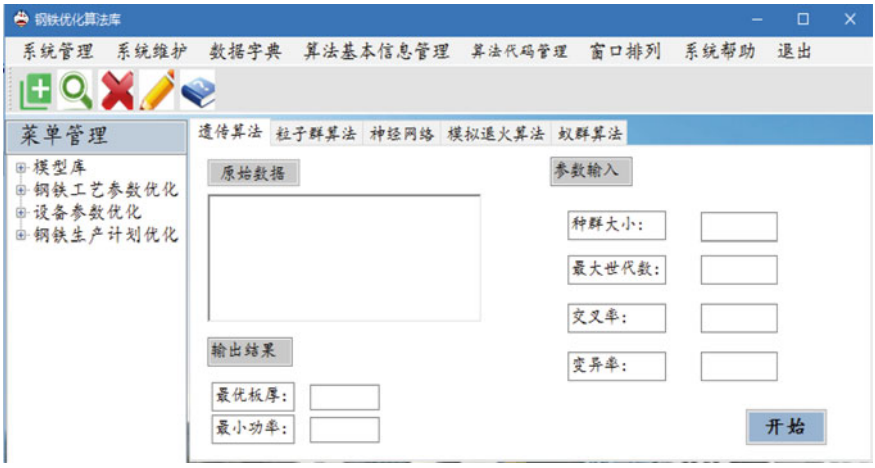
4.2 Knowledge Base System Interface Display

The knowledge base system is divided into three parts: First of all, the user login interface, entering the user name password all correct can enter the system. Next, the register form, the registered user name and password will be written to the backend database, so that the user name and password are correct when logged in. At last is the main form, the user selects the optimization algorithm according to the steel optimization direction that they need. The interface is shown as follows (Figs. 8 and 9).

As shown in Fig. 10, The goal of algorithmic knowledge base is to solve the problems of various constraints and complicated process in the process of steel production. The optimal scheduling of resources, energy and technology is the core, and the optimization problem of production is solved quickly based on various

Fig. 8 Login form

Fig. 9 Register form



Main form

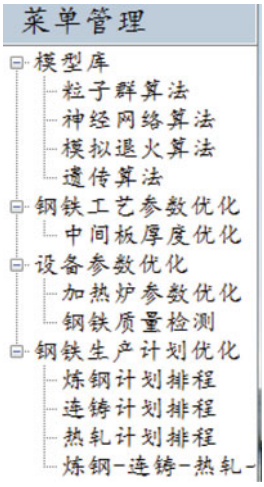


Fig. 10 System directory tree

intelligent optimization algorithms. The knowledge base is divided into two main parts. The first part is the model base, which contains a variety of intelligent optimization algorithms, which mainly include Particle Swarm Optimization algorithm, neural network algorithm, simulated annealing algorithm and genetic algorithm. For individual problem managers, the algorithm can be added to the system at any time. The second part is the application of intelligent optimization algorithm for steel production process. Users can choose suitable application direction according to specific problems.

5 Conclusions

This paper is based on the optimization problem in the steel production process and summarized the optimization type of the steel production process. A knowledge base system of steel production process optimization algorithm is established. The commonly used optimization algorithm and the optimization of steel production process are provided. It is convenient for enterprises to optimize the problems in the production process and improve the production efficiency and quality of the enterprises.

Acknowledgements The authors would like to express appreciations to mentors in Shanghai University and Shanghai Baosight Software Corporation for their valuable comments and other helps. Thanks for the China's National science and technology pillar program's funding. The program number is No. 2015BAF22B01.

References

1. Hu Y (2006) Development and application of decision support system. China Water Conservancy and Hydropower Press, Beijing
2. Chen X (2000) Principle and application of decision support system. Tsinghua University Press, Beijing
3. Yan C (2006) Practical software engineering course. China Water Conservancy and Hydropower Press, Beijing
4. Lu Z, Zhai Q, Xie G, et al (1997) Energy consumption forecast of steel industry in China. J Steel 5:69, 74
5. Zhang G (2013) Research and application of energy intensity index system of steel enterprises. Northeastern University
6. Wei B, Zeng Z et al (1992) The research on the energy standard of the energy standard. J. China Energy 2:24–26
7. Zhang Y (2004) SQL server 2000 database programming. Mechanical Industry Press, Beijing
8. Korczak JJ, Maciaszek LA, Stafford GJ (1989) Knowledge base for database design[C]. International symposium on database systems for advanced applications. Seoul, Korea, April. DBLP. pp 61–68

Study on the Friction Characteristics of Two Polytetrafluoroethylene (PTFE) Coatings

Jian Wu, Te Li and Lanzhong Guo

Abstract PTFE is a kind of good self-lubricating materials because of its excellent lubricating performance. In this paper, the friction experiment is conducted to study the tribological properties of sintering and bonding coatings of PTFE. The results show that the friction coefficient of the two coatings increases firstly and then decreases with the increasing in the relative speed, and increases slowly with the increase of load. The bonding coatings of PTFE has shown good lubricating properties in low speed for its large thickness, and sintering coatings has shown good lubricating properties under high load for its strong adhesion strength.

Keywords PTFE · Preparation processes · Coefficient of friction
Wear appearance

1 Introduction

Polytetrafluoroethylene (PTFE) is an ideal self-lubricating material which has very low friction coefficient, high chemical and thermal stability, better electrical insulating property, and excellent resistance to high and low temperature [1, 2]. The commodity use involves aerospace, automotive electronics, petrochemical, construction, textile and other industrial sectors. The PTFE become one of the important and indispensable material.

The main wear mechanism of PTFE is adhesion wear, fatigue wear and abrasive wear [3, 4], besides, there are also aging wear, creep wear and chemical wear

J. Wu (✉) · T. Li · L. Guo
Department of Mechanical Engineering,
Changshu Institute of Technology, Changshu, China
e-mail: wujian123abc@163.com

J. Wu · T. Li · L. Guo
Jiangsu Key Construction Laboratory of Elevator Intelligent Safety, Jiangsu, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_58

between the contact surface [5]. To improve the wear resistance of PTFE, some functional particles were used to modify the tribological properties, such as carbon fiber, rare earths, carbon nanotube and so on [6–8].

However, in industrial applications, it cannot be used as anti-wear material alone because of its poor mechanical properties, bad thermal conductivity and creep resistance, high wear rate, and Low bearing strength. Therefore, PTFE is used as coating to prepare on the surface of substrate, so that it has a certain strength, and it can be used in special places. In this article, two kings coating have been introduced, and the friction characteristics have been experimental researched under velocity and load.

2 Materials and Preparations

The PTFE coating is a membrane mixed with PTFE and filings, and covered in metal matrix by different processes, which is a main application form in industry. PTFE coating manufacture are mainly containing sintering and bonding.

2.1 The Sintering Process

The structure of sintering coating as shown in Fig. 1, which including three coating: steel backing, copper layer and PTFE. To get a high load strength, the steel backing has used as the base, the PTFE has placed in the outside as the lubricating layer, and the copper layer between the steel backing with PTFE to fix the PTFE.

The process of sintering coating is shown in Fig. 2.

(a) The steel tape has been polished, and the copper power was laid on it with a certain thickness, At a high temperature (1000 °C), the copper powder is sintered with the base material to form a physical base layer.

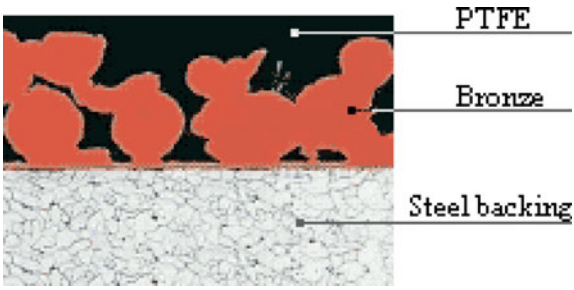


Fig. 1 The structure of sintering coating

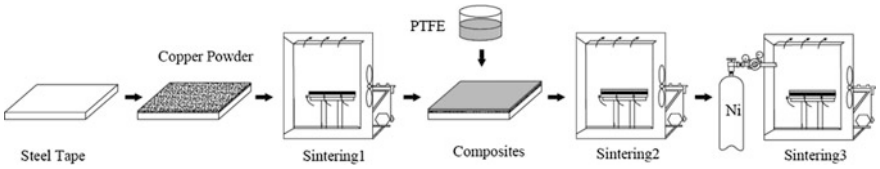


Fig. 2 The process of sintering coating

(b) The rational dispatching schemes of PTFE coating can be obtained by Chemical Crosslinking of Polyphenylene sulfide (PPS). The mixture is mixed in a blender, and be sifted by 100 mesh sieves. Put a certain adhesive and harder, the rheid coating has been prepared at 15–25 °C stir for 5–10 min.
(c) The rheid coating is evenly rolled on the base material, with the thickness is 100 μm.
(d) To solidify the surface coating, the evenly coated material in an oven for high temperature baking, the baking temperature is 220–280 °C and the baking time is about 30 min.
(e) To crystallize the coating completely, the solidified PTFE coating is sintered in a nitrogen furnace. The sintering temperature is 370–390 °C and the time is about 40 min.

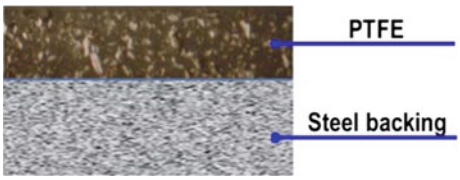
2.2 The Bonding Process

The structure of bonding coating as shown Fig. 3. It contains only two layers: steel backing and PTFE layer. The PTFE thin film is attached to the substrate through special adhesive technology.

The process of bonding coating is shown in Fig. 4.

(a) The rational dispatching schemes of PTFE coating can be obtained by Chemical Crosslinking of Polyphenylene sulfide (PPS). The mixture is mixed in a blender, and be sifted by 100 mesh sieves. A cylindrical PTFE body has been prepared when the PTFE powder is pressed by molding process.
(b) To crystallize the coating completely, the cylindrical PTFE body is sintered in a nitrogen furnace. The sintering temperature is 370–390 °C and the time is about 2 h.

Fig. 3 The structure of bonding coating



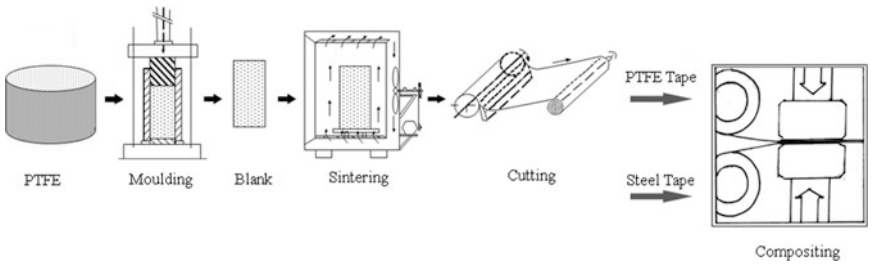


Fig. 4 The process of bonding coating

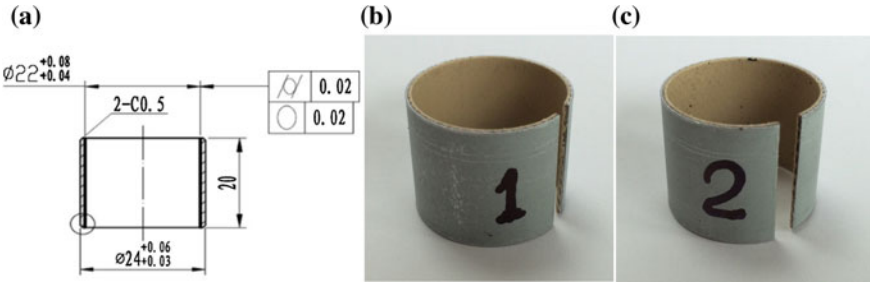


Fig. 5 The picture of two samples **a** the parts drawing; **b** the bonding sample; **c** the sintering sample

(c) The crystalline the coating is cutting with 250 μm film thickness.
(d) The polished base layer and the processed PTFE film are bonded to PTFE coating using a press machine.

2.3 Preparation for Samples

According to the above two kinds of preparation, The drawing and pictures of sample as shown in Fig. 5. To meet the requirements of experimental, the specimens were manufactured as cylindrical.

3 Experiment

The whole test equipment as shown in Fig. 6. A motor provides a rotational movement. The output axis connections a sensor which measure the torque, and friction part is placed at the right of the picture. Friction pair is mainly made up of

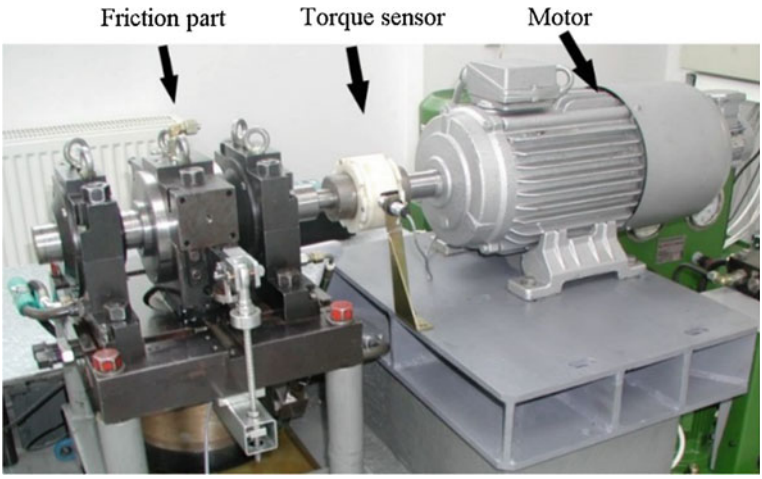


Fig. 6 The picture of experiment equipment

center of rotation axil (45#) and PTFE samples, which assembled with 0.02 mm. When the axil is running, the friction is happened in sliding pair, the torque sensor measured the torque which can transforming into the coefficient of friction.

In those test, the PTFE coating friction coefficient is discussed under the condition of loads and speeds. And the wear surfaces are characterized by using SEM. The experiments were conducted in the ambient temperature is 25 °C and the relative humidity is 80%.

4 Results and Discussion

4.1 Coefficient of Friction

(1) The friction coefficient varies with the speed under the constant load

In this experiment, the parameters are as follows. The preload is 2 MPa, the relative sliding speed is 0.2–2.0 m/s, speed interval is 0.2 m/s. The experiment ran for 10 min at every speed and took the mean of the friction coefficient. Figure 7 is shown the friction coefficient of two PTFE coating under different speed.

When the sliding speed is lower, the friction coefficient of the sintering process PTFE coating is higher than that of the boding coating. At the speed of 0.8–1.6 m/s, the friction coefficient of the two coatings has emerge a reducing trend, and the bonding coatings of PTFE decreases more. At the speed of 1.6–2.0 m/s, the friction coefficient of the two coatings increases with the increase of relative speed, and the sintering coatings of PTFE increase more.

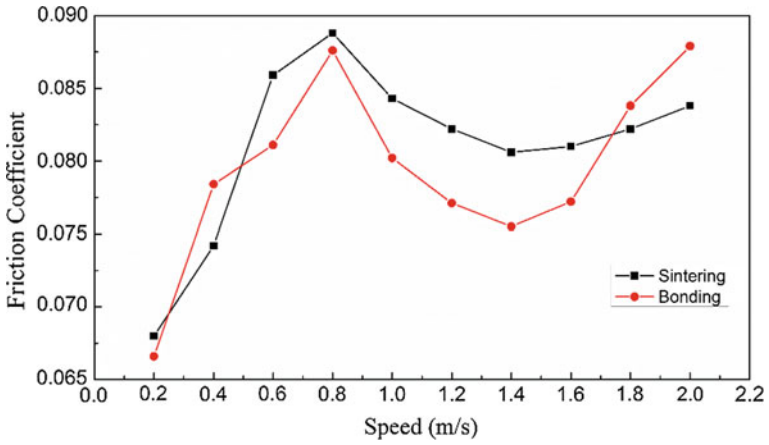


Fig. 7 The curve of the friction coefficient changing with speed

(2) The friction coefficient varies with the load in the constant speed

The speed is chosen at 1.0 m/s according to experiment above, whose friction coefficient is relative stable and high. The load is 0.25–2.5 MPa, load interval is 0.25 MPa. The experiment ran for 10 min at every load and took the mean of the friction coefficient. Figure 8 is shown the friction coefficient of two PTFE coating under different load.

The two friction coefficient values showed a slow increasing trend. The friction coefficient of the bonding coating increased with the increase of the load, and gradually became stable, with a larger value of 0.07. When the load exceeded

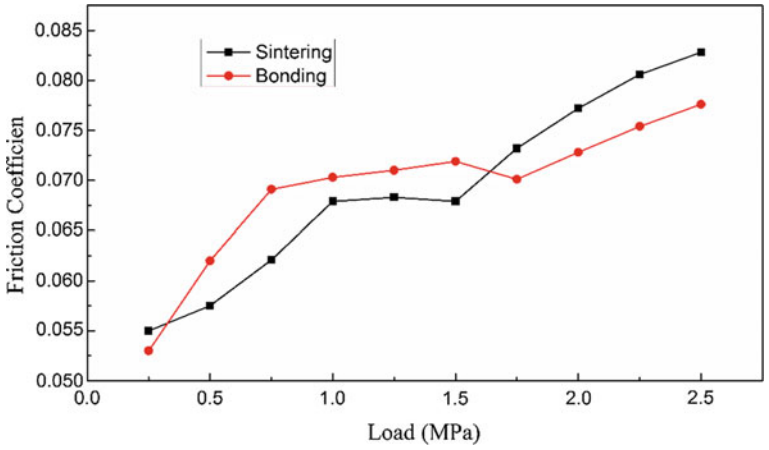


Fig. 8 The curve of the friction coefficient changing with load

1.75 MPa, it continued to increase. The friction coefficient of sintering coating gradually increased in the initial stage, the change was greater than that of the bonding coating. When the applied load exceeded 0.75 MPa, the friction coefficient tended to be stable and slightly decreased, and tended to be stable. The friction coefficient of the sintering coating increased as the load exceeded 1.5 MPa, and the friction coefficient was higher than that of bonding coating.

4.2 Analysis

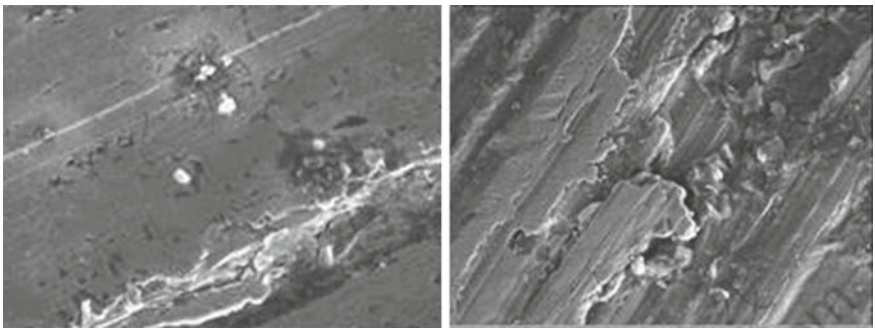
(1) The influence of speed on wear

Figure 9 shows the wear appearance of two coating with 0.8 m/s (1.0 MPa). The sintering coating has good adhesion performance with the substrate at low speed, which turns to be lower friction coefficient and the wear appearance is shown in Fig. 9a.

Figure 10 shows the wear appearance of two coating with 1.6 m/s. In Fig. 10a, as the coating was thin, the coating came off during the friction at high relative speed, the substrate material directly involved in the wear, which turns to be higher friction coefficient. In Fig. 10b, the abrasion was not very clear, and some materials were piled up in the abrasion. The wearing PTFE coating was involved in the friction surface. At higher speeds, the friction coefficient was relatively low.

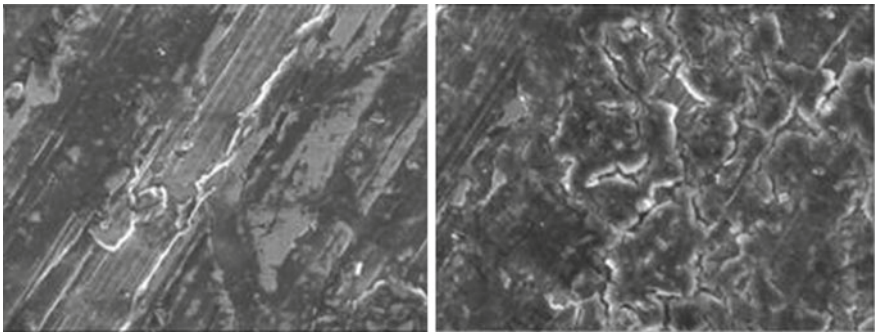
(2) Effects of different loads on wear

Figure 8 shows the friction coefficient of bonding coatings was obvious lower than that of sintering coatings. Figure 11 showed the wear appearance of two coating under 1.75 MPa (0.8 m/s). Figure 11a indicated that surface wear was more obvious on sintering coatings. The wear appearance in Fig. 11b was obscure, material stocking was more than that of sintering coatings, which showed great



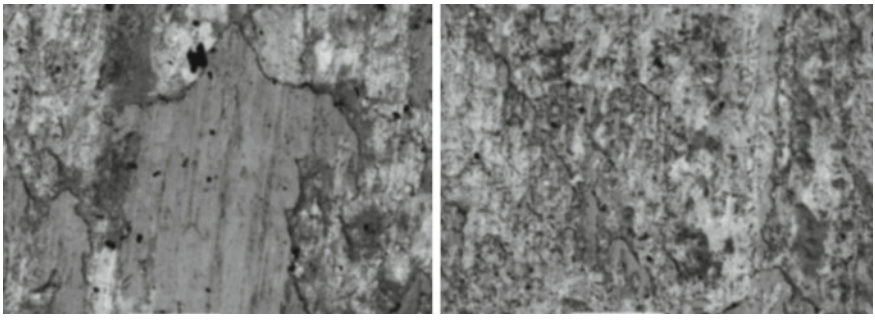
(a) Sintering coating (b) Bonding coating

Fig. 9 The wear appearance of two coating with 0.8 m/s



(a) Sintering coating (b) Bonding coating

Fig. 10 The wear appearance of two coating with 1.6 m/s



(a) Sintering coating (b) Bonding coating

Fig. 11 The wear appearance of two coating under 1.75 MPa

lubricating property. The PTFE debris produced during the wear process had some wear reduction characteristics, so the friction coefficient of the sintering coating rise rapidly and was larger than that of the bonding coating.

5 Conclusion

The tribological properties of sintering and bonding coatings of PTFE show obvious difference at different speeds and load. For sintering coatings, PTFE is bonded to the substrate surface by sintering, which shows good adhesive capacity. As a result, the friction coefficient of sintering coatings of PTEE is low under low load and speed. When the load and speed is large, the thin coating are come off in the friction, the substrate material is directly involved in friction, and the friction coefficient becomes larger. Due to poor adhesion ability, the friction coefficient of bonding coatings shows a great change under low load and speed. When the load

and speed is large, the PTFE coatings help to reduce friction, and friction coefficient of is lower than that of sintered coatings of PTFE. The PTFE debris is produced during the wear which further relieves wear.

Acknowledgements This project is supported by Doctoral Science Foundation of Changshu Institute of Technology (No. KYZ2015054Z).

References

1. Bely VA et al (1982) Friction and wear in polymer-based materials. Pergamon Press, Oxford
2. Blanchet TA, Peng YL (1998) Wear resistant irradiated FEP/unirradiated PTFE composites. Wear 214(2):186–191
3. He P et al (2008) Tribological properties and wear mechanism of ekonol composites. Polym Mater Sci Eng 24(8):86–89
4. Huang C et al (2006) Friction and wear properties of inorganic filled Ekonol/PTFE composites. Acta Materiae Compositae Sinica 23(4):36–40
5. Ren LQ et al (2006) Experimental investigation of bionic rough curved soil cutting blade surface to reduce soil adhesion and friction. Soil Tillage Res 85(1–2):1–12
6. Zhang Z, Liu W, Xue Q (2010) Effects of various kinds of fillers on the tribological behavior of polytetrafluoroethylene composites under dry and oil-lubricated conditions. J Appl Polym Sci 80(11):1891–1897
7. Arsic D (2016) Tribological behavior of four types of filler metals for hard facing under dry conditions. Ind Lubr Tribol 68:35–38
8. Wang H et al (2007) Effects of fibrous fillers on friction and wear properties of polytetrafluoroethylene composites under dry or wet conditions. Particuology 5(6):414–419

A Hybrid Nested Partitions Method for Bi-objective Job Shop Scheduling Problem Considering Energy Consumption and Makespan

Mei Dai, Xin Li and Limin Mao

Abstract The energy consumption in discrete manufacturing draws extensive attention as the global warming and air pollution. It is critical to consider energy consumption as well as traditional optimization objective of job shop scheduling. This paper addresses a bi-objective optimization problem for job shop scheduling and proposes a hybrid nested partition algorithm with NSGA-II. The nested partition method is a partitioning- and sampling-based strategy that focuses computational effort on the most promising region of the solution space while maintaining a global perspective. The sampling strategy of NSGA-II samples from each of the sub-regions and from the aggregated surrounding region with Pareto-optimal solutions, so that the algorithm quickly converges. The proposed method is evaluated in an application case and the results show that it really optimize the energy consumption as well as makspan.

Keywords Nested partition · NSGA-II · Job shop scheduling · Energy consumption

1 Introduction

As is well-known, excessive carbon dioxide emissions lead to the global warming and air pollution, however, saving energy will help to reduce pollution significantly. The manufacturing industry has always been an intensive energy user. Manufacturing enterprises equipmented with machine tool are responsible for approximately one-half of the total global energy consumption, 38% of CO_2 emissions, and significant pollution [1]. Hence reducing the machining system energy was deemed to be one of the strategies to improve sustainability in manufacturing.

M. Dai (✉) · X. Li · L. Mao
School of Electrical Engineering and Automation, Changshu Institute of Technology, No. 99, 3rd South Ring Road, Changshu 215500, Jiangsu, China
e-mail: daidaimei_0_0@hotmail.com

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_59

Scheduling is one of the most important sub-systems in a manufacturing system. Previous studies concentrated more on traditional criterions such as productivity and rarely consider energy consumption issues. In fact, scheduling can significantly impact the energy consumption of the entire manufacturing system.

For decades, there has been an upswing in studies on energy-aware scheduling. Mouzon presented dispatching rules to address the problem of scheduling on a single machine which minimized the idle energy consumption [2, 3]. Furthermore, some research work addressed the energy-aware optimization in the field of flow shops. Herrmann reported that scheduling lot size on two production lines can improve energy efficiency for machining the same quantity of parts [4]. Min Dai introduced an energy-efficient model for flexible flow shop scheduling [1]. Fang K presents a new model of the flow shop scheduling problem that considers peak power load and energy consumption in addition to cycle time [5]. Chen analyzed the effective scheduling of machine startup and shutdown energy consumption [6]. Liu proposed a multi-objective scheduling method with reducing energy consumption as one of the objectives [7]. Yin proposed an effective multi-objective genetic algorithm based on simplex lattice design [8].

The optimization method mainly focused on mathematical programming method which are suitable for small scale scheduling and intelligence optimization method such as simulated annealing, tabu search, genetic algorithm etc. Although the intelligence optimization method were good at large scale combined optimization problem, but the computation was rather complex. Nested partition was proposed to address combined optimization problem by Shi [9]. This method were effectively utilized to solve scheduling problem, TSP and feature selection problem. Wu proposed a hybrid nested partition algorithm with genetic algorithm for scheduling in job shop problem [10]. Pervious studies utilizing NP rarely focused on multi-objective. This paper proposes a hybrid nested partitions method for bi-objective job shop scheduling considering energy consumption and makespan.

The rest of the paper are organized as follows: the mathematical model of this problem is formulated considering the energy consumption as well as makespan in Sect. 2. In Sect. 3, hybrid nested partitions algorithm for solving the mathematical model is presented. Section 4 illustrates an application case to verify the proposed method. Finally, concluding remarks are outlined.

2 Mathematical Model

2.1 Optimization Objective of Makespan

A n $\times$ m job shop scheduling problem is described that n jobs need to be processed on m machines, the processing sequence of jobs is defined in advance, and each operation of a job has its own processing time. Each machine can only process one job at the same time; each operation cannot be interrupted. The job shop scheduling

problem can then be formulated as the following mixed integer problem. The decision variables are defined as follows.

$$a_{ihk} = \begin{cases} 1 & \text{if machine } h \text{ processes job } i \text{ before machine } h; \\ 0 & \text{else.} \end{cases} \tag{1}$$

$$x_{ijk} = \begin{cases} 1 & \text{if machine } k \text{ processes job } i \text{ before job } j; \\ 0 & \text{else.} \end{cases} \tag{2}$$

Let h, k denote the serial number of machine; i, j denote the serial number of job; c_{ik} denotes the completed time of job i on machine k; p_{ik} denotes the processing time on machine k; M is a considerable large positive number. The optimization objective of makespan can be formulated with constrains as follow.

$$\text{Min} \max_{1 \le k \le m} \left\{ \max_{1 \le i \le n} c_{ik} \right\} \tag{3}$$

Subject to

$$c_{ik} - p_{ik} + M(1 - a_{ihk}) \ge c_{ih} \tag{4}$$

$$c_{jk} - c_{ik} + M(1 - x_{ijk}) \ge p_{jk} \tag{5}$$

$$c_{ik} \ge 0 \tag{6}$$

where

$$i, j = 1, 2, \ldots, n$$

$$h, k = 1, 2, \ldots, m$$

2.2 Energy Criterion for the Problem

According to the research by Yan [11], the total machining energy E consists of both E_m, the operations energy, and E_w, the idle energy of the machine tools.

$$E = E_w + E_m \tag{7}$$

The idle energy of machine tools E_w can be calculated by multiplying the machine idle power, P_o, by the idle time before the next operation, t_w. [2, 3].

$$E_w = P_o \times t_w \tag{8}$$

The machining energy, E_m, consists of idle energy during job setup (E_s) and cutting energy (E_c).

$$E_m = E_s + E_c \tag{9}$$

E_s is calculated by multiplying P_o by the tool and work piece setup time, t_s. For multiple-function machine tools, energy consumption in switching time t_g is also considered as shown in Eq. (11):

$$E_s = P_o \times t_s + P_o \times t_g \tag{10}$$

$$E_c = (P_O + kv) \times t_C \tag{11}$$

where v is the material removal rate, t_c is the cutting time(s). The processing time is the sum of t_c, t_s, and t_g. k is the specific cutting energy, which can be obtained by cutting experiments or handbook with empirical data for different materials.

2.3 Mathematical Model

A mixed integer programming mathematical model can be established for the energy consumption objective and makespan objective for job shop scheduling problem. The optimization functions of the job shop scheduling problem are described as follows.

$$\text{Min}\, f_1 = \max_{1 \le k \le m} \left\{ \max_{1 \le i \le n} c_{ik} \right\} \tag{12}$$

$$\text{Min}\ f_2 = E = \sum_{i=1}^{n} \sum_{k=1}^{m} E_{ik} + \sum_{k=1}^{m} P_{ok} t_{wk} \tag{13}$$

Constraints

$$c_{ik} - p_{ik} + M(1 - a_{ihk}) \ge c_{ih}$$

$$c_{jk} - c_{ik} + M(1 - x_{ijk}) \ge p_{jk}$$

$$c_{ik} \ge 0$$

$$p_{ik} = t_{sk} + t_{ck}$$

3 Scheduling Algorithm

The optimization problems of job shop scheduling are larger-scale discrete optimization problems. The Nested Partitions (NP) for optimization has been proven to be a useful approach for effectively solving large-scale discrete optimization problems [9]. The NP algorithm consists of four parts, that are partitioning, sampling, select promising index and backtracking. The framework of the proposed method for bi-objective optimization of job shop scheduling problem is simply illustrated in Fig. 1.

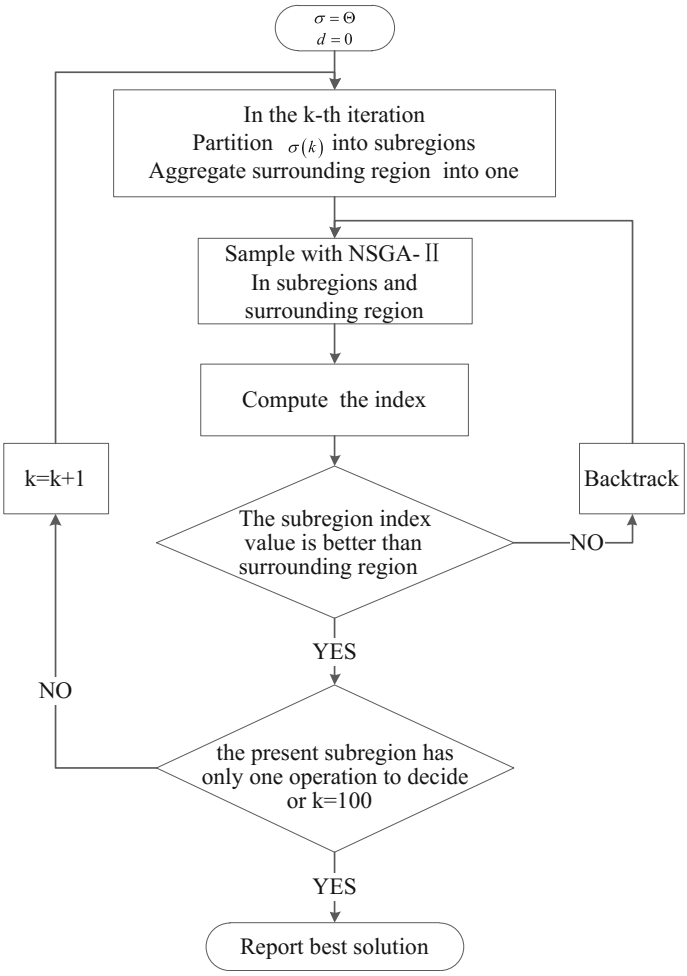


Fig. 1 Framework of the hybrid nested partition algorithm

3.1 Partitioning Scheme

Partitioning is the first key step for the NP approach. For a n × m job shop scheduling problem, a schedule $\theta \in \Theta$ can be written as $\theta = (O) = \{o_1, o_2, \ldots, o_{n\times m}\}$. In this paper, the schedule θ uses a permutation with m-repetitions of job numbers. Each job number occurs m times in the schedule θ. The partitioning strategy is to assign one operation at each level of the partitioning tree. If the most promising region occurs in each iteration that means backtracking is not necessary, then the minimum iteration times is n × m.

3.2 Sampling Scheme

The next step is to sample from each of the subregions and from the aggregated surrounding region. Assume that a region $\sigma \in \sum$ is being sampled, then $d(\sigma)$ operations have already been assigned and $n - d(\sigma)$ operations have not been fixed. In this paper, we use the Non-dominated Sorting Genetic Algorithm to sample. The steps of the NSGA-II to solve the model are as follows:

Step 1. Initial parent population P_0 with the size N is generated randomly, and the fast non-dominated sorting is implemented on P_0. The genes of the individuals before the k-th gene are fixed by current partition.

Step 2. Individual with each rank is sorted again based on the density based clustering strategy to evaluate the population density.

Step 3. Individuals selected are implemented by the binary tournament operator in the mating pool. And between two individuals, the selected individual is the one with the lower rank. If two individuals are on the same level, the winner is the one with the larger value in the crowding distance.

Step 4. The offspring population Q_t is generated by the genetic operations such as crossover and mutation where t denotes the number of generations in the mating pool. In this paper, the crossover and mutation operators are defined referring to Tang [12].

Step 5. An integrated population P_t is created by combining P_t and Q_t, and fitness values are designed to all individuals by the non-dominated sorting and crowded distance sorting.

Step 6. Finally, individuals with better fitness are selected by elitist sorting and these become the parent individuals P_{t+1}.

Step 7. Steps 2–6 are repeated until $t = t_{max}$.

Step 8. Individuals with rank 1 among parents at P_{max} are Pareto-optimal solutions.

3.3 Select Promising Index Scheme

The promising index of the algorithm is the best sampled found at each subregion and agrees with the original performance function on regions of maximum depth. The promising index function is defined based on the optimizing objectives of the model. In each iteration, the promising index region is determined with the region of the best solution. In this paper, the adapted weighted sum method is used to construct the promising index function. Considering our mathematical model in Sect. 2.3, makespan f_1 and energy consumption f_2 are not incompatible objective function. Hence, multiple objective function f_1 and f_2 are combined into one overall objective function to evaluate each sample in Pareto-optimal solutions set. The value for evaluating the promising index function is calculated as follow:

$$f = w_1 \frac{f_1 - f_{1\min}}{f_{1\max} - f_{1\min}} + w_2 \frac{f_2 - f_{2\min}}{f_{2\max} - f_{2\min}} \tag{14}$$

Where w_1, w_2 are the weight coefficients depending on decision-making.

3.4 Scheduling Backtracking Scheme

If the promising index value equals to the surrounding region, then backtrack to a large region containing the current most promising region. The backtracking strategy plays a critical role in the method that good backtracking strategy can reduce the iteration times. In this paper, this step backtracks to the super region of the current most promising region.

4 Case Study

A 4 × 5 job shop scheduling application case is given to illustrate the hybrid Nested Partitions method. As shown in Table 1, the experiment data of the application case consists of the operation sequence of each job, the setup time and cutting time of each machine and the energy consumption of each machine on specific operation. Table 2 gives the idle power of machine tools and energy consumption in alternative machines.

The related parameters of the NSGA-II algorithm are set as follows: the population size is 30, the maximum evolution generation is 50, the crossover rate is 100% and the mutation rate is 10%. The encoding schema used in this paper is the operation-based representation encoding as well as partitioning tree.

As shown in Table 3, the 4 × 5 job shop scheduling result presents the total energy consumption including machining energy and idle energy is optimized as

Table 1 Experiment data of the application case

Job	Operation sequence	t_s/t_c(min)/Energy consumption of machines(10^{-3} KWh)
1	{O_{12}, O_{14}, O_{13}}	{1.5/5.6, 3.0/3.7, 2.5/5.0}/{228.7, 97.2, 310.4}
2	{O_{22}, O_{25}, O_{23}, O_{24}}	{1.0/0.3, 2.5/1.2, 2.5/0.5, 3.0/1.6}/{40.2, 29.5, 97.5, 50.6}
3	{O_{31}, O_{34}, O_{32}, O_{33}, O_{35}}	{1.5/1.3, 2.5/3.0, 1.0/0.9, 4.0/1.2, 4.0/1.2}/{60.4, 66.0, 57.0, 173.5, 89.7}
4	{O_{44}, O_{41}, O_{45}}	{3.0/9.6, 3.0/12.3, 6.5/3.0}/{195.3, 367.9, 94.1}

Table 2 Idle power of machine tools

Machine	M_1	M_2	M_3	M_4	M_5
Idle power (W)	995	1485	1910	600	430

Table 3 Schedule solution of the proposed method

Machine	Operation time [start time-finish time (min)]
M_1	O_{31}(9.8–12.6) O_{41}(12.6–28.2)
M_2	O_{22}(9.7–11.0) O_{12}(11.0–18.1) O_{32}(18.1–20.0)
M_3	O_{33}(20.0–25.7) O_{23}(25.7–28.7) O_{13}(24.8–32.3)
M_4	O_{44}(0–12.6) O_{34}(12.6–18.1) O_{14}(18.1–24.8) O_{24}(24.8–29.4)
M_5	O_{25}(11.0–14.7) O_{35}(25.7–30.7) O_{45}(30.7–40.2)

well as makspan for the case. The total operation time is 40.2 min while the total energy consumption of machines amount to 1677.91 Wh, where only 4.5% is wasted for idle running of machining tools.

5 Conclusion

This paper addresses an energy-aware job shop scheduling problem to optimize the machining system energy consumption at the job shop level. The Nested Partition algorithm quickly converges by concentrating the search on these sub-regions. As the two objectives of energy and makspan are on tardness, the NSGA-II algorithm effectively sample the sub-region. The proposed hybrid method combines the advantages of Nested Partition algorithm and NSGA-II, which improves the search efficiency.

The authors of this paper are planning to improve the partition strategy and promising index calculation to enhance the performances of the proposed algorithm in the future. Besides, integrating energy consumption into dynamic scheduling is a challenging research work.

References

1. Dai M, Tang D, Giret A, Salido MA et al (2013) Energy-efficient scheduling for a flexible flow shop using an improved genetic-simulated annealing algorithm. Robot Comput Integr Manufact 29:418–429
2. Mouzon G, Yildirim MB, Twomey J (2007) Operational methods for minimization of energy consumption of manufacturing equipment. Int J Prod Res 45(18–19):4247–4271
3. Mouzon GC, Yildirim MB (2008) A framework to minimize total energy consumption and total tardiness on a single machine. Int J Sustain Eng 1(2):105–116
4. Herrmann C, Thiede S (2009) Process chain simulation to foster energy efficiency in manufacturing. CIRP J Manufact Sci Technol 1(4):221–229
5. Fang K, Uhan N, Zhao F, Sutherland WJ (2011) A new approach to scheduling in manufacturing for power consumption and carbon footprint reduction. J Manufact Sys 30:234–240
6. Chen GR, Liang L, Arinez J, Biller S (2013) Energy-efficient production systems through schedule-based operations. IEEE Trans Autom Sci Eng 10(1):27–37
7. Liu Y, Dong H, Lohse N, Petrovic S, Gindy N (2014) An investigation into minimizing total energy consumption and total weighted tardiness in job shops. J Clean Prod 65:87–96
8. Yin L, Li X, Gao L, Lu C, Zhang Z (2017) Energy-efficient job shop scheduling problem with variable spindle speed using a novel multi-objective algorithm. Adv Mech Eng 9(4):1–21
9. Shi L (2000) Nested partitions method for global optimization. Oper Res 48:390–407
10. Wu W, Wei J, Guan X (2010) Proceedings of the 2009 IEEE international conference on robotics and biomimetics, Guilin, China, 19–23 Dec 2009
11. Yan H, Li Y, Wu T, Sutherland JW (2015) An energy-responsive optimization method for machine tool selection and operation sequence in flexible machining job shops. J Clean Prod 87(2015):245–254
12. Tang Q, He M, He X, Zhang L, Floudas CA (2015) Robust optimization scheduling of flexible job shops under stochastic processing times. Comput Integr Manufact Sys 21 (4):1002–1012

Research on the Framework of Quality Prediction in Intelligent Manufacturing

Min Ji

Abstract Product quality is the lifeblood of manufacturing enterprises. Under the background of intelligent manufacture, quality prediction is affected by lots of intricate elements. Furthermore, the customization level of products is one of the important factors which can support the enterprises in the increasingly fierce competition. Therefore, the high customization level means that it is impossible to collect a large quantity of quality data because of multi variety and small batch production mode. This paper proposes the principle and the framework of quality prediction in intelligent manufacturing. The framework adopts the theory of combinational prediction because it is more stable, more reliable and more accurate than single prediction model. The research of this paper will supply a new prediction method for the manufacturers and improve the accuracy of quality prediction and reduces the production risk.

Keywords Single prediction model selection · Combinational prediction model TOPSIS · Support vector regression · Quality prediction

1 Introduction

Accurate quality prediction not only allows the manufacturing enterprises to under-stand the quality of products in advance, but also allows them to be well-prepared deal with the potential risks. In most cases, the enterprises select a variety of single prediction models to predict the quality based on different assumptions. According to the results of the hypothesis testing, they select the best single prediction model from all of the models. Because each single prediction model explains and analyzes the historical data from different angles, only part of

M. Ji (✉)
School of Intelligent Manufacturing and Control Engineering, College of Engineering, Shanghai Polytechnic University, No. 2360, Jinhai Road, Shanghai, China
e-mail: jimin@sspu.edu.cn

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_60

the information are extracted from the historical data by each single prediction model. Thus, the prediction accuracy is affected. Obviously, this prediction process is not suitable in the rapidly changing manufacture industry [4].

Different single prediction models have their own advantages and disadvantages. Each model contains different information of historical data, which are interrelated and complementary to each other. Therefore, it will be more scientific if we combine the results of several single prediction models appropriately to make full use of the useful information contained in each single prediction model. This is the main concept of the combinational prediction, which can improve the precision and reduce the risk of prediction [1, 2].

The idea of combinatorial prediction proposed by Bates and Granger [3] in 1969 has received extensive attentions from scholars. In recent years, the theory of combinatorial prediction has been widely used in various fields, such as market demand, grain yield, meteorological prediction, power load and exchange rate. Several different single prediction models predict the same problem, and then the combinatorial prediction model will combine the predicted value of each single model. The key of combinatorial prediction is to utilize comprehensively the information of each single model, and to improve the prediction accuracy as much as possible [7].

However, while the theory of combinatorial prediction is booming, it is confronted with two major problems in the practical application.

1. Among multiple single prediction models, which models can be selected as the in-put of the combinational prediction model?
2. With the development of manufacture industry, the product quality is influenced by the more and more complicated external and internal factors. Which combinational prediction model can not only adapt to the complex environment, but also can predict more accurately and improve the generalization ability?

Under this background, we have carried on the research and the discussion to the above questions. Based on the previous researches, this paper proposes the framework of quality prediction, which integrates the selection arithmetic of single prediction model and combinational prediction model, which is especially suitable for the production mode of multi variety and small batch.

2 The Framework of Combinational Prediction Model in Intelligent Manufacturing

In this section, we will analyze the characteristics of intelligent manufacture firstly. According to the characteristics, we propose a framework of combinational prediction to predict the product quality in intelligent manufacturing.

2.1 The Characteristics of Intelligent Manufacturing

With the development of manufacturing technology, enterprises change from extensive mode of unitary-breed, huge-batch, rigid production to multi-varieties and small batch customized production, which has become the main production mode in intelligent manufacturing.

In the mode of multi-varieties and small batch customized production, the products have strong special purpose, the production processes are complex, parameters are diversified, the processing methods are not uniform, the environmental conditions of production are changeable and uncertain, production planning and control are cumbersome, the supervision is difficult to implement and the production management is dynamic [8]. The above characteristics will increase enterprise inventory and production cost and reduce the quality of the products.

Based on the above analysis, the principle of TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) method is proposed to serve as the selection model of single prediction models, which can evaluate the single prediction models synthetically. Furthermore, support vector machine Support vector regression machine is a small sample learning method [6]. It has good generalization ability even in the case of a small number of training samples. Especially, it is very effective to deal with non-linear regression problem. Considering the intelligent manufacturing is multivarieties and small batch customized production, we think support vector regression machine is suitable for quality prediction.

2.2 The Selection of Single Prediction Model

TOPSIS method is a sorting method of approximating the ideal solution, which is proposed by C. L. Hwang and K. Yoon [5]. The basic concept is to establish the initial decision matrix and the evaluation index system, and then based on the normalized initial matrix, a virtual ideal solution and a virtual negative solution are constructed. The ideal solution is a virtual solution composed of the optimal values of each evaluation index, and the negative solution is composed of the worst values of each evaluation index. The vertical distance between the evaluation objects (the optional single prediction models) and the ideal solution and the negative solution can be calculated. If an evaluation object is not only close to the ideal solution but also far away from the negative solution, this evaluation object has good performance. The algorithm is easy to apply and can evaluate the pros and cons of each single prediction model systematically and comprehensively. The specific process (Fig. 1) is as follows.

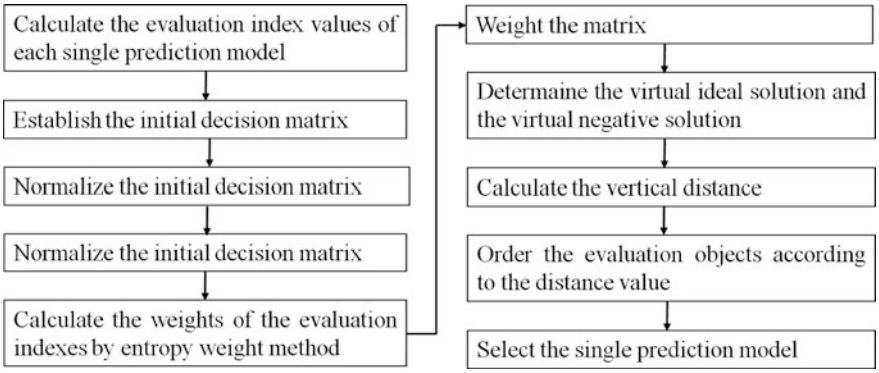


Fig. 1 The selection process of single prediction models

2.3 The Framework of Support Vector Regression Combinational Prediction Model

At present, there is no unique and universally accepted method and criterion about how to construct non-linear combinational prediction model. Several models are popular, such as polynomial function, nonlinear programming model and neural network. However, only when the size of the training samples is large, the effectiveness of these algorithms can be presented, which doesn't adapt to the multi-varieties and small batch customized production mode in intelligent manufacturing.

After the research on the support vector regression machine, we put forward the framework of the combinational prediction model based on support vector regression machine, and divide the modeling process into four stages.

The first two stages are related to the selection of the single prediction models, and the last two stages construct a nonlinear combinational prediction model for the single prediction models selected in the first two stages by using the support vector regression. This proposed framework solves the problem of the single prediction model selection. It selects several single prediction models which has good performances in each evaluation index, and lays a solid foundation for the successful implementation of the combinational prediction. And particularly in the complex intelligent production environment, the support vector regression model can be used to self-learn the sample data effectively, and it has strong generalization ability, especially for the non-linear prediction with small sample. The framework is presented in Fig. 2.

The first stage is the pre-processing phase. At first, it is necessary to determine the prediction object and collect the historical data. According to the characteristics of the prediction object, we can select several single prediction models which

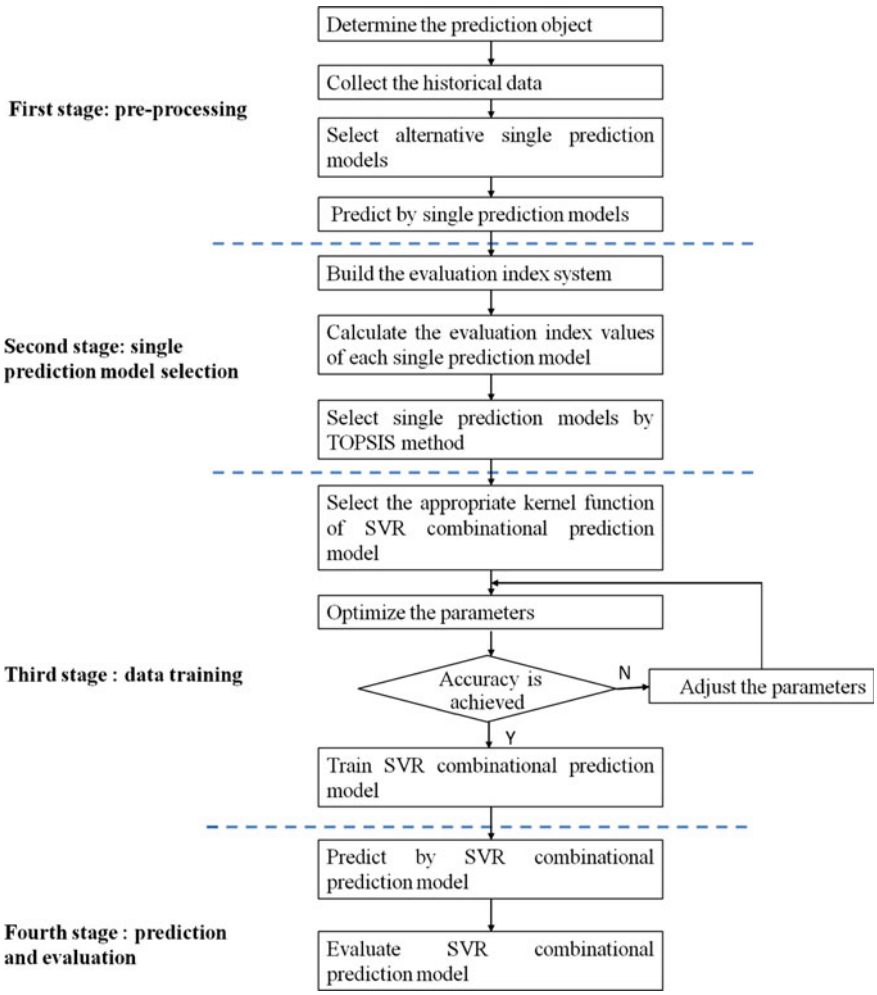


Fig. 2 The framework of SVR combination prediction Model

satisfy the trend of the prediction object and determine the parameters of each single prediction model. The predictive value of each single prediction model is obtained based on the historical data.

The second stage is the single prediction model selection. After obtaining the prediction values of single models, we can constructs the evaluation index system and calculate the index values of each single prediction model. Then, the single prediction models are selected by using the TOPSIS method. According to the actual demand and the evaluation index system, the single prediction models which have good performances are chosen as the inputs of the combinational prediction model.

The third stage is the data training phrase. The predictive value of each single model is divided into two groups: training sample and prediction sample. The training sample is the input vector of the combinational prediction model based on the support vector regression machine, and the corresponding actual observation values are used as the output of the combinational prediction model. Next step is to select the appropriate kernel function and use the cross verification method based on grid search to optimize the parameters. Finally, the optimal parameters are put into the support vector regression combinational prediction model. After the model is trained, a support vector regression model is obtained.

The fourth stage is the prediction and evaluation phase. We can predict the pre-diction samples using the support vector regression combinational prediction model which is generated in the third stage. After obtaining the combinational predictive value, we can compare it with the actual observation value to evaluate the model performance.

3 Conclusion

In this paper, we focus on the quality prediction in intelligent manufacturing. TOPSIS method is proposed to select the single prediction model. It sets up a comprehensive scientific evaluation system to evaluate the performance of single prediction models from many aspects. Based on the statistical learning theory, a framework of the combinational prediction model based on support vector regression machine is proposed, which has good generalization and robustness. This framework can transform the low dimension nonlinear space into high-dimensional linear space, which overcomes the phenomenon of "over learning" and "dimensionality disaster". It is very suitable for multi-varieties and small batch customized production mode in intelligent manufacturing.

Acknowledgements This work is supported by the foundation of Shanghai Polytechnic University (Grant No. EGD17XQD12).

References

1. Bates JM, Granger CWJ (1969) The combination of forecasts. Oper Res Q 20(4):451–468
2. Chen H (2008) The validity principle and application of combinational prediction. Science Press, Beijing
3. Diebold XF (2003) Element of forecasting. South-Western, Div of Thomson Learning
4. Elliott G, Granger CWJ, Timmermann AG (2006) Handbook of economic forecasting. Elsevier, Amsterdam
5. Vapnik VN (1995) The nature of statistical learning theory. Springer, Berlin

6. Wang J (2012) The research and application on several combinational prediction models. Ningxia University, China
7. Zhou Y (2009) The research on the TOPSIS method in multi-attribute decision-making. Wuhan University of Technology, China
8. Zheng WW, Liang JJ (2007) Research on the quality integration control of multi-varieties and small volume production process. Mach Des Manufact Eng 36(7):19–21

Deep Learning Approach to Multiple Features Sequence Analysis in Predictive Maintenance

Jin Yuan, Kesheng Wang and Yi Wang

Abstract The representation learning of life cycle dataset has its particularity in the correlation of different features and the dependency of adjacent sampling time. This paper addresses the difficulty of segmentation to high-dimensional nonlinear life cycle long CBM data, and propose a new deep learning approach based on unsupervised representation learning named Autoencoder for rolling bearing diagnosis. Two kinds of Autoencoder with encoder and decoder model are developed respectively using fully connected and convolutional hidden layers to automatically extract the dataset's representative features. Compared to the fully connected one, the convolutional Autoencoder shows clearer in a lower dimensional feature space by preserving the local neighborhood structure, and more effective to discover subjectively the intrinsic structure of nonlinear high-dimensional data of deterioration process.

Keywords Predictive maintenance · Multiple features sequence
Bearing life cycle · Deep learning · Autoencoder

J. Yuan
College of Mechanical and Electronic Engineering,
Shandong Agricultural University, Tai'an, China
e-mail: jyuan@sdau.edu.cn; jin.yuan@ntnu.no

J. Yuan · K. Wang (✉)
Department of Mechanical and Industrial Engineering, NTNU, Trondheim, Norway
e-mail: kesheng.wang@ntnu.no

Y. Wang
The School of Business, Plymouth University, Plymouth, UK
e-mail: yi.wang@plymouth.ac.uk

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_61

1 Introduction

With the rapid development of automation, informationization and intelligent technology, the modern production equipment structure is becoming more and more complicated. The maintenance theory and method of the production system cannot be applied to the modern multi-device complex manufacturing system, mainly in the variable structures and working principles leaded to different evolution patterns of equipment health, which show the diversity of non-bathtub curve underwent strong randomness. Practice shows that frequent regular maintenance and static long-term planning cannot improve the reliability of equipment, thus convectional maintenance planning due to incapable dynamic prediction may easily fall into lack or excess of maintenance. That is, many expensive components are replaced prematurely or deteriorative parts still in service, even lead to accidents, which results in the lack of optimization of system maintenance.

The health prediction of the equipment refers to the use of accurate and effective model to carry out the state detection, symptom diagnosis and failure prediction of the equipment health recession process, and predict the future health deterioration trend by data mining and analysis of the current running state of the equipment. Predictive maintenance decision-making key technologies include the following aspects: data acquisition and monitoring technology, data preprocessing and features extraction method, health statuses classification and deterioration trend prediction algorithm, equipment maintenance planning model, system scheduling optimization strategy [1].

The life cycle of the equipment is different, the performance and status of the equipment is also different, the periodic maintenance measures should be different. At the same time, the degree of damage to the fault is different, the failure to take the maintenance measures should also be different. Therefore, it is necessary to introduce the concept of life cycle to qualitatively distinguish and describe the age and health status of the equipment. It is necessary to establish an evaluation model to evaluate the life cycle of the equipment.

On the other hand, Bearings are the most widely used mechanical parts. Due to natural degradation process and various reasons during operation, even in normal circumstances, the wear and tear of bearings that fatigues lead to premature damage. Because the tests and validation of fault diagnostic and prognostic methods are easy to introduce and to perform on the bearing test rig, furthermore, the long and slow degradation of bearing could to be acquired with monitoring signal sequences. Therefore, to study Condition-Based Maintenance (CBM) based bearings is an important research direction in predictive maintenance and health management field. Thus, we using this kind of dataset to test our algorithm in this paper.

Vibration signals are often contaminated by noise and thus unusable to distinguish direct from different machine statuses and fault types. Feature extraction can increase the signal to noise ratio to improve the correct identification by find some intrinsical ingredients. Several vibration of rotating machinery fault diagnosis techniques have been applied to. Some conventional time-domain, frequent-domain

and time-frequent domain's feature extraction by various signal processing algorithms to represent the manifold structure in vibration data space [2, 3].

The essence of deep learning [4] is to automatically learn more meaningful features of high dimensional complex data by constructing machine learning model and layer-by-layer feature transformation with multiple hidden layers, and discover the equipment inherent deterioration patterns, and ultimately improve the prediction accuracy. The deep Autoencoder network [5] is a mainstream representation learning model and high dimensional visualization tool, by studying the deep nonlinear network structure of the input, reveals the distributed characteristics of the input data, in addition to its ability to learn essential features from a number of samples, these excellent performance is far from the traditional shallow model is far from comparable. It can better model the rolling bearing vibration data and optimize the structure of continuous deep Autoencoder network model to further improve the prediction accuracy and effectively complete the fault prediction of rolling bearing.

This paper addresses the difficulty of handling high-dimensional nonlinear life cycle long CBM data, we propose a new deep learning approach based on unsupervised representation learning named Autoencoder for rolling bearing faults diagnosis. The rest of this paper is organized as follows: Section 2 illustrates the methodology of multiple features sequence analysis in predictive maintenance for the segmentation problem. Two kind of Autoencoder with encoder and decoder model is addressed respectively using fully connected and convolutional hidden layers to handle automatically extract the dataset's representative features. The experiments and discussion of the proposed solution is evaluated in Sect. 3. Finally, the conclusions are given in Sect. 4.

2 Methodology of Multiple Features Sequence Analysis in Predictive Maintenance

The evaluation of the life cycle of the equipment is the baseline of this paper approach, which is a basis for further determining the hazard level of a failure mode in the monitoring system of equipment, also as a foundation for determining maintenance implement. According to the fault occurrence, the life cycle of equipment conventionally is divided into three stages: run-in period (early failure period), maturity stage (occasional fault period) and recession period (loss failure period). Different maintenance strategies are used for these three stages in the traditional preventive maintenance. However, with the equipment failure rate and performance changes more complex and diverse, to address more reasonable and comprehensive qualitative description of equipment status and performance, it is very necessary and useful to make more detailed stage division in equipment predictive maintenance.

2.1 Architecture of Multiple Features Sequence Analysis and Segmentation

Data collected from the progressive deteriorated equipment often gather together, more complicated is that those data has time dependent, thus it is far difficult to segment naturally into clusters, where the characteristics of equipment status in the same cluster are similar and the characteristics of equipment status in different clusters are dissimilar. Therefore, the aim of this paper is to enhance the representative visualization of monitored multiple time series sequences leveraged by deep learning, and to give out a more reasonable segmentation.

Figure 1 shows the system overview of Autoencoder method for multiple features sequence analysis and segmentation in predictive maintenance. Our model consists of five main parts: condition based monitoring based data collection, conventional feature extraction, representation learning with Autoencoder, dimensionality reduced data visualization and life cycle status segmentation. Given some multiple sequences of multi-channels sensors or one main sensor signal (A), and then those multi-features could be extracted by time domain analysis, frequency domain analysis and time-frequency domain (B). The slices of multi-correlative time series could be generated, and the slices from different features are stacked together. Then, they are be feed to two kind of Autoencoder with encoder and decoder net components to train by representation learning process and multiple features reconstruct process (C). A dimensionality reduced latent feature representation are effectively predicted by the encoders, and then leveraged by T-SNE or PCA etc. visualization tools to show the intrinsical manifold structure mapping to 2D plane (D). From the distribution and the deterioration of points represented to every sampling statuses, using proper clustering algorithm, combined with subjective decided cluster number, to determine the corresponding cluster category representing current status in life cycle (E).

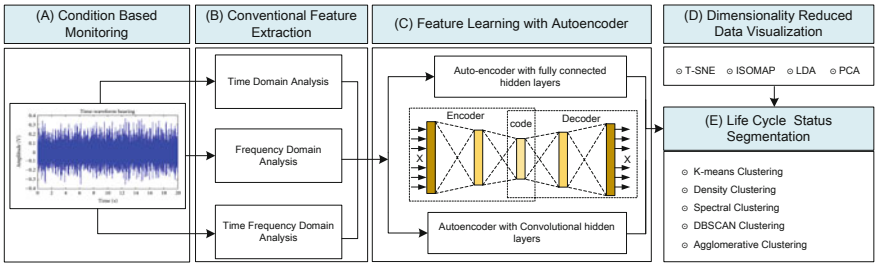


Fig. 1 System overview of Autoencoder method for multiple features sequence analysis and segmentation in predictive maintenance

2.2 Two Autoencoder Network Structures

The purpose of dimensionality reduction is to obtain a more compact representation of the original high-dimensional data, a representation that nonetheless captures all the information necessary for higher-level decision-making. We propose two types deep encoder-decoder network for dimensionality reduce of multiple features sequence. Figure 2 shows the novel cascade convolution hidden layers and MaxPooling2D hidden layers formed the Encoder network structure, and cascade convolution hidden layers and UpSampling2D hidden layers formed the Decoder network structure, which together formed Autoencoder to be fed with multi-feature sequences X and output the reconstructed X', and the Encoder could output the encoded and reduced low dimensional manifold, which is utilized to effectively discover the structure of this manifold by having the network to reconstruct the input despite some constraint and the Encoder model the spatial and sequential correlations of sequence slices with step size.

For the performance comparative, we also utilized the fully connected hidden layers to handle the same multi-feature sequences X, but without using slice of sequences. Both two types deep encoder network of Autoencoder could automatically extract the dataset's representative features.

2.3 Visualization and Segmentation

The deep Autoencoder successfully re-extracted the reduced representations by identified the patterns of vibration in every life cycle statuses. Visualization of high-dimensional data is a challenging problem of fault diagnosis and prognosis to deal with the high-dimensional and nonlinear data collected from the complete

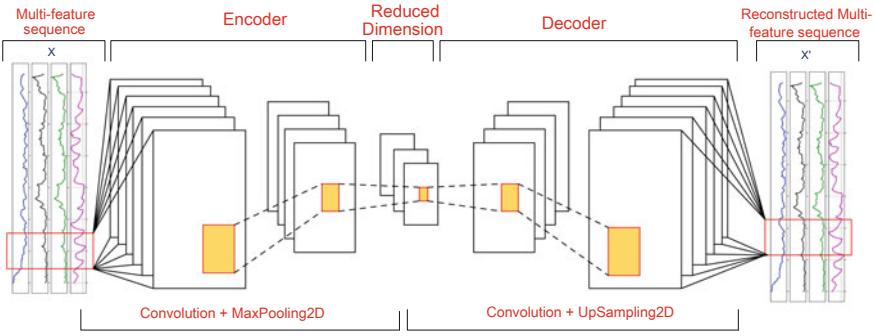


Fig. 2 The structure of Convolutional Autoencoder with encoder subnetwork and decoder subnetwork to learn the process of multi-sequence compression and reconstruction

information of operating machinery. The t-SNE algorithm [6] outperforms existing state-of-the-art techniques for visualizing a variety of real-world datasets.

Both Autoencoder and t-SNE try to find a lower dimensionality embedding of multi-features sequence. However, the Autoencoder tries to minimize the reconstruction error, while t-SNE tries to find a lower dimensional space and at the same time it tries to preserve the neighborhood distances. t-SNE is usually preferred for visualizations.

There is no uniform evaluation Index for clustering performance, unless there are labels of cluster category, because the objective functions of different clustering algorithms are very different. The segmentation of life cycle dataset has its particularity in the correlation of different features and the dependency of adjacent sampling time, therefore the clustering algorithm should be propagation based.

3 Experiments and Discussion

We conduct experiments on one bearing life cycle dataset to show the utilization of this approach. The convolution layers Autoencoders are compared with the traditional fully connected layers methods that stack the correlation of different features or different signal from channels. We evaluate our multiple features sequence learning scheme on typical bearing datasets to verify our method for modeling the temporal dependency.

3.1 Bearing Life Cycle Dataset

The bearing life cycle dataset is captured from an experimentation platform named PRONOSTIA [7] shown in Fig. 3. The database provides real experimental data that characterize the degradation of ball bearings along the whole operational life (until their total failure). One of the dataset life cycle had the operating conditions: 1800 rpm and 4000 N, and the dataset had sampled every 10 s with 1/10 s sampling times with sampling frequency 25.6 kHz. That is, the life cycle dataset used in this experiments has 2803 bins data (whole raw signal shown in Fig. 4).

3.2 Feature Extraction and Representation Learning with Autoencoder

In this Experiment, 12 dimensional classical features are extracted in time domain: Mean, Maximum Amplitude, Root Mean Square, Peak-to-Peak, Pulse Index, Crest

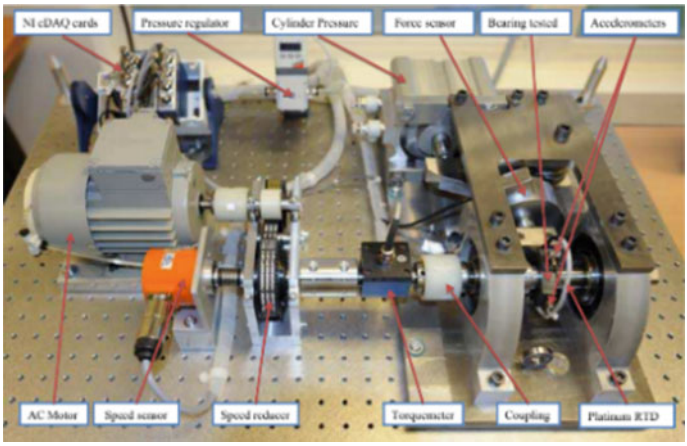
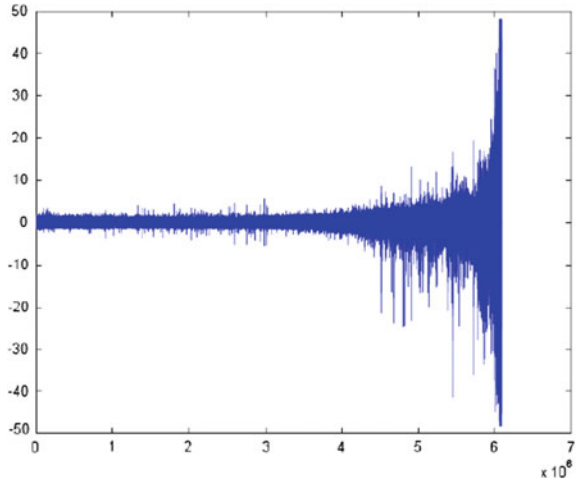


Fig. 3 Overview of PRONOSTIA

Fig. 4 A vibration raw signal



Factor, Waveform Factor, Standard Variance, Skewness, Kurtosis, Entropy, and Energy. Partly features is shown in Fig. 5.

3.3 Visualization and Segmentation

Figure 6a shows the deterioration process utilized by the color mapping its sampling points in bearing life cycle, which the representation of multi-feature sequences is learned by the fully connected Autoencoder and mapped it to 2D space by t-SNE. Figure 6b shows the same deterioration process, but the representation of

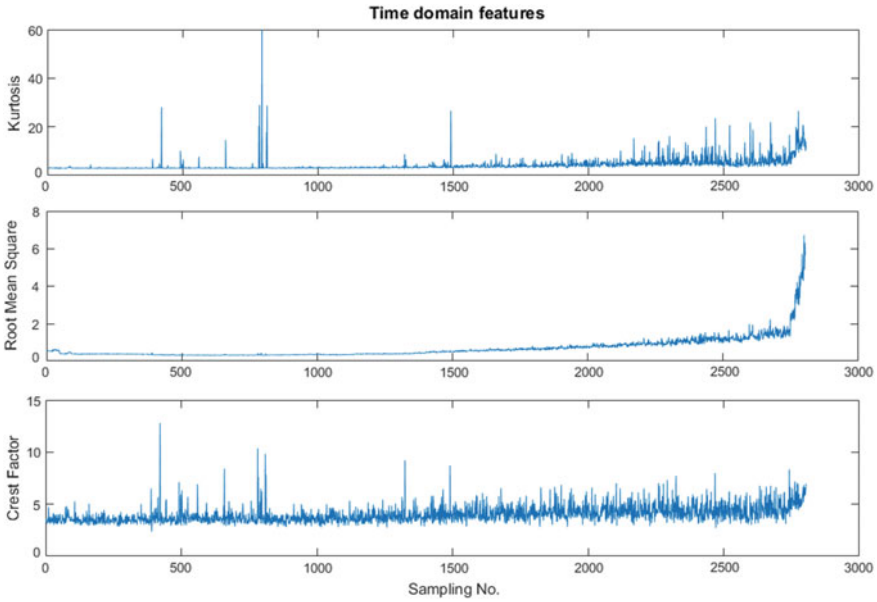


Fig. 5 Part of the features sequence of bearing life cycle extracted from time domain

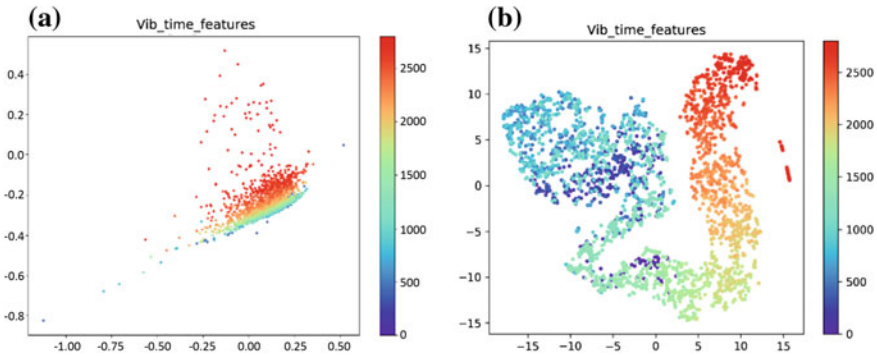


Fig. 6 Visualization of deterioration process of bearing life cycle dataset in 2D space. **a** Fully connected Autoencoder + t-SNE. **b** Convolutional Autoencoder + t-SNE

multi-feature sequences is learned by the convolutional Autoencoder and mapped it to 2D space by t-SNE with same parameters. Compared to the fully connected one, the convolutional Autoencoder shows clearer in a lower dimensional feature space by preserving the local neighborhood structure, and more effective to discover subjectively the intrinsic structure of nonlinear high-dimensional data of deterioration process.

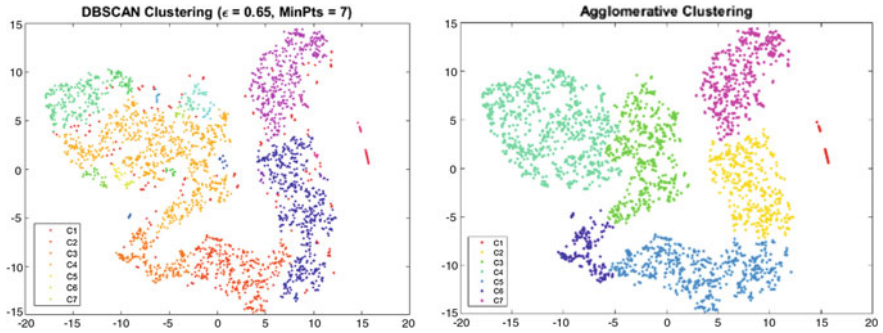


Fig. 7 Segmentation of deterioration process of bearing life cycle dataset with 7 clusters respectively by DBSCAN and agglomerative clustering

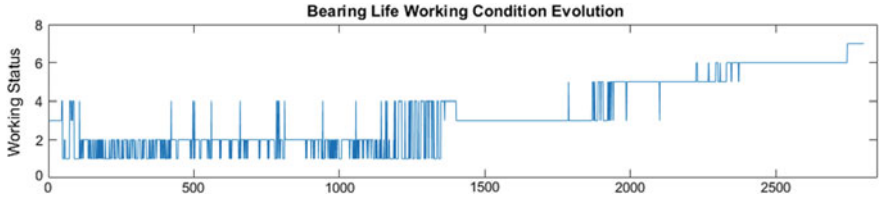


Fig. 8 Status changing in the deterioration process of bearing life cycle dataset

Based on the result of visualization, we decide the life cycle dataset should be segmented to 7 clusters. We compared Hierarchical-based named agglomerative clustering, Density-based DBSCAN clustering and Distance-based K-means clustering to investigate grouping the life cycle bearing dataset [8]. The latter two clustering algorithms have big problem of partitioning observations according to locations and distances from each other, that cannot distinguish the variable of different stage in the deterioration process, for example, C7 the next to failure status at the end of sequence has some points in the early status belonged to same cluster in the DBSCAN clustering in Fig. 7, while this phenomenon is not shown in the right plot. The Agglomerative clustering has clearer centroid clusters of representative characterization. Finally, mapping the clustering category to the status changing (Fig. 8) in the deterioration process of bearing life cycle dataset, which is more interpretable for the decision making and more useful in predictive maintenance.

4 Conclusions

This paper addresses the difficulty of handling high-dimensional nonlinear life cycle CBM data, we propose a new deep learning approach based on unsupervised representation learning named Autoencoder for bearing fault diagnosis. Two kinds

of Autoencoder with encoder and decoder model are developed respectively using fully connected and convolutional hidden layers to handle automatically extract the dataset's representative features. Moreover, the achieved features also are visualized with state-of-the-art dimensionality reduction and clustering methods. The experiments show that convolutional Autoencoder has more feasibility and effectiveness at preserving the local neighborhood structure, and discovering the intrinsic structure of nonlinear high-dimensional data for the representation learning of multi-feature sequences.

Acknowledgements The work is part of project MonitorX. Which is supported by Norwegian Research Council (NFR).

References

1. Wang KS (2014) Key techniques in intelligent predictive maintenance (IPdM)—a framework of intelligent faults diagnosis and prognosis system (IFDaPS). Adv Mater Res 1039:490–505
2. Mobley RK (2003) An introduction to predictive maintenance, 2nd edn. Butterworth Heinemann, Boston, USA
3. Grall A, Dieulle L, Berenguer C, Roussignol M (2002) Continuous-time predictive-maintenance scheduling for a deteriorating system. IEEE Trans Reliab 51(2):141–150
4. Goodfellow IJ, Le QV, Saxe AM et al (2009) Measuring invariances in deep networks Adv Neural Inf Proc Syst 22:646–654
5. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313:504
6. Maaten LVD, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9(Nov): 2579–2605
7. van der Maaten L, Hinton GE (2008) Visualizing high-dimensional data using t-SNE. J Mach Learn Res 9:2579–2605
8. Patrick N, Rafael G, Kamal M et al (2012) PRONOSTIA: an experimental platform for bearings accelerated life test. In: IEEE international conference on prognostics and health management

Sorting System of Robot Based on Vision Detection

Qin Qin, Dongdong Zhu, Zimei Tu and Jianxiong Hong

Abstract According to industrial assembly line sorting activities, the sorting system of industrial robot was designed based on vision detection. This system uses industrial camera to obtain workpiece images and corresponding QR code label information. The workpiece recognition method is used to identify the workpiece and calculate the center coordinates of the workpiece which are unified to the robot coordinate system and communicate with the robot. The robot can accurately grasp and sort the workpiece. The system uses LabVIEW to achieve image acquisition, image processing and coordinate transformation tasks. Using MOTOCOM32.dll which provided by Yaskawa company to establish real-time communication with robot controller. The system has a high running speed and accuracy, and can be applied to automatic sorting operations in the industrial assembly line.

Keywords Industrial robot · Machine vision · Automatic sorting
LabVIEW

1 Introduction

With the development of manufacturing industry, visual inspection technology has found more and more applications in industrial production, and plays an essential role in improvement of production efficiency and intellectualization of production [1, 2]. The work of industrial robots is inseparable from sorting. Sorting jobs are a critical link in most assembly lines. Sorting of workpieces is also an integral part of the link in batch production of industrial products. On conventional assembly lines, most of sorting jobs are performed manually. The continuous and repeated actions of sorting performed by operators can no longer meet the needs of production, and have greatly lowered work efficiency and accuracy.

Q. Qin (✉) · D. Zhu · Z. Tu · J. Hong
Shanghai Polytechnic University, 2360 Jin Hai Road,
Pudong District, Shanghai 201209, China
e-mail: qinqin@sspu.edu.cn

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_62

2 System Hardware Design

The system hardware is mainly composed of industrial robot, pneumatic equipment, industrial camera and lens, light source, computer, router and sorting object (small plastic pieces of different shapes). The connection of main system hardware is shown in the Fig. 1:

3 Software Design

In this system, LabVIEW is used as the software platform, which includes machine visual image processing package NI Vision Pack and general-purpose industrial camera driver package NI-IMAQdx, to write and design host computer programs and perform the jobs of image acquisition, analysis and recognition processing. The robot programming is done on the robot teaching-programming pendant.

According to the expected system implementation, the work flow can be divided into seven major sub-flows as shown in the flowchart below (Fig. 2):

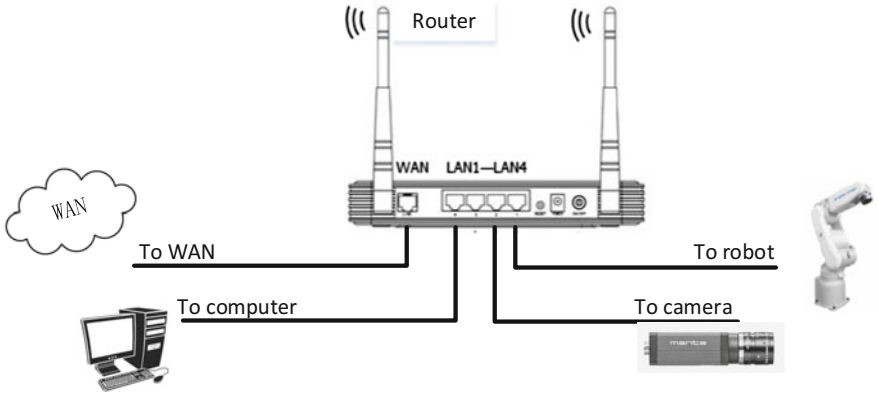
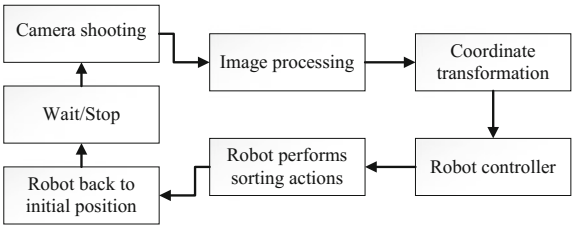


Fig. 1 Hardware connection schematic

Fig. 2 Work flowchart



Software design contains the design of host computer control programs and robot actuation programs. The functions of host computer programs include control of industrial camera shooting, image processing, target location, coordinate transformation and coordinate information transmission. The system can display images and system work status in a real-time mode, and also has the function of system start/stop.

The functions of robot actuation programs include receiving start/stop signals, and controlling robot's performance of sorting actions in a real-time mode with the workpiece coordinate information and target location information sent from the host computer.

3.1 Host Computer Program Design

3.1.1 Program Flow Design

The state machine is chosen as the program design pattern in combination with the operation flow of the system. Communication between the program and the robot is realized by calling MOTOCOM32.dll (MOTOCOM32 is a software development tool package provided by Yaskawa Electric Co., Ltd. of Japan for its MOTOMAN series robots, which contains such functions as data file transmission, robot control, and I/O signal write/read, for the users to develop their application programs for their own needs). On the basis of the system's operation flow, the control program is written in the design pattern of the state machine as shown in Fig. 3.

3.1.2 Image Acquisition, Processing and Coordinate Transformation

General-purpose industrial camera driver package NI-IMAQdx provided by NI Company is used to perform various operations of the camera. First, IMAQ Set Calibration Info VI used for image correction, and then for image recognition and coordinate transformation. The program in Fig. 4 shows the use of IMAQ Count Objects VI to recognize the position of workpiece in image, and application of IMAQ Read QR Code VI to read QR code information of the relevant workpiece.

The coordinates of a workpiece in an image are represented in pixel points. They must be unified with the coordinates of the robot for actuation. The robot provides the users with the right to define the coordinates by themselves, therefore, axes x and y in the user coordinate system of the robot are so defined as to align them with axes u and v in the coordinates in the image, axis z is perpendicular to axes x and y, and the origin is noted as O, as shown in Fig. 5a.

The purpose can thus be achieved by transformation of coordinates. Here, the coordinates of the workpiece in the image coordinate system are assumed to be (u, v), the origin of the image coordinate system to coincide with that of the user coordinate system, and note as O, and the visual field shot with the camera to be a

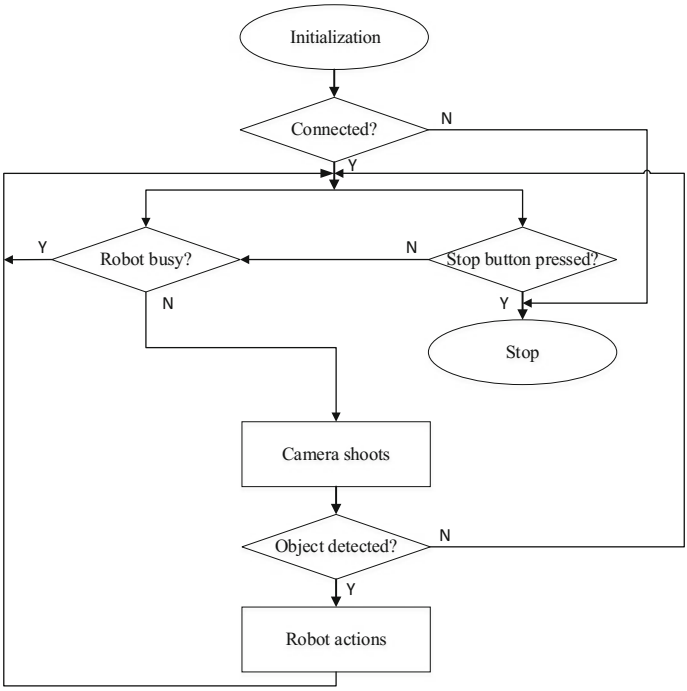


Fig. 3 Program flowchart

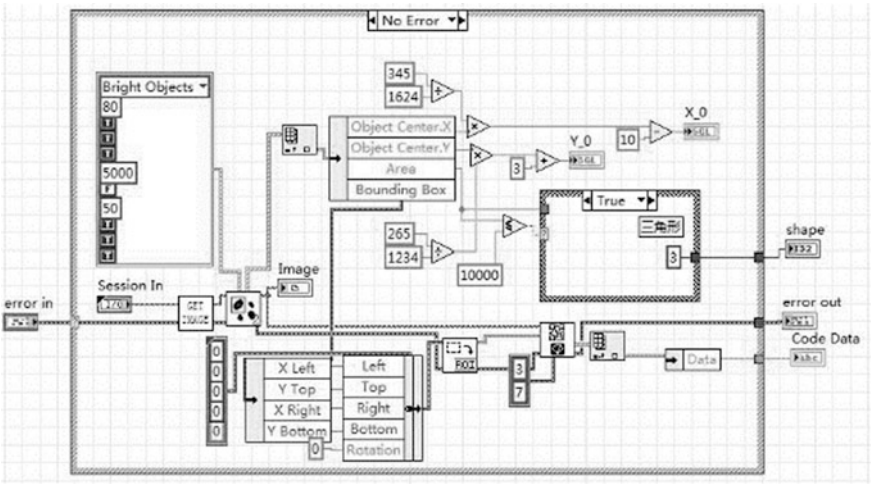


Fig. 4 Image processing and coordinate transformation procedure

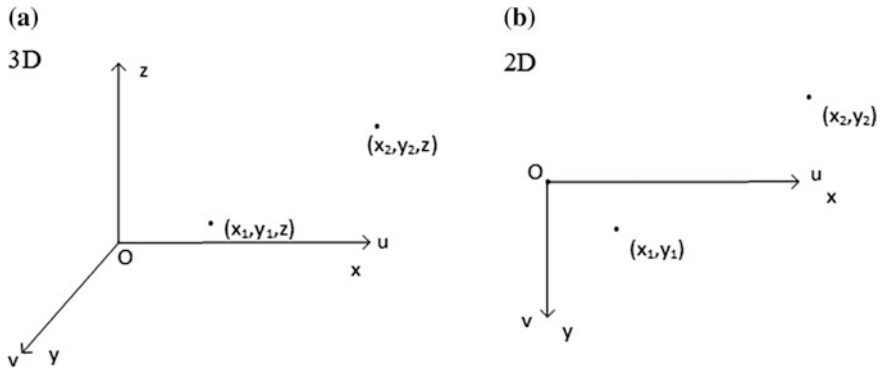


Fig. 5 Coordinate system

rectangular area of 345 mm × 265 mm, and the effective pixels of the camera are known to be 1624 * 1234, the coordinates (x_1, y_1) of the workpiece center in the user coordinate system expressed in the unit of mm can be obtained:

$$x_1 = \mathrm{u} \times 345/1624 \tag{1}$$

$$y_1 = \mathrm{v} \times 265/1234 \tag{2}$$

Take the coordinates of the robot end point in the robot user coordinate system as (x_2, y_2) as shown in Fig. 5b. After transformation of coordinates, the distances the robot needs to move in the directions of x and y are:

$$x = |x_1 - x_2| \tag{3}$$

$$y = |y_1 - y_2| \tag{4}$$

As for axis z, because the vertical distance between robot actuating point and workpiece center point is fixed, the operation distance from the actuating point to the center point on axis z needs to be designated in robot programming only.

3.2 Robot Actuation Program Design

The function of the host computer control program is to pick up and process images and to transform coordinates and design the flow state machine. In this system, the robot needs to actuate along the predetermined route and then comes back to the initial position. The operation flow is shown in the chart below (Fig. 6):

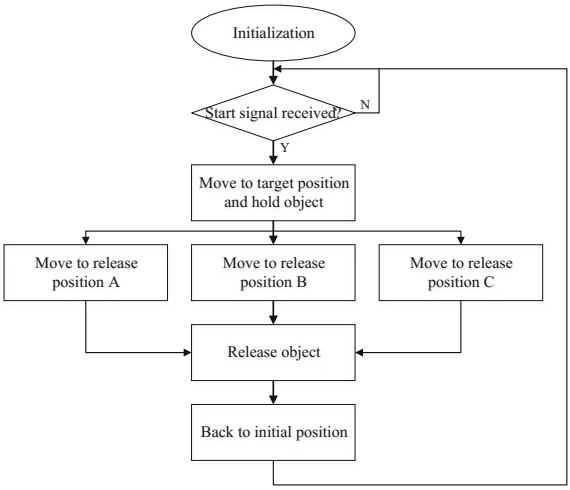


Fig. 6 Flowchart of robot actuation

4 System Operation Effect

In this experimental system, triangular, square and circular workpieces are used as the sorting objects. The interface can display the shape of the workpiece currently recognized and count, and simultaneously read the QR code information on the workpiece and store it. The communication state indicates communication with the robot controller. The robot operation state indicator shows whether the robot is busy or not.

After the host computer calls the start task function, the robot starts to actuate sorting actions. The output controls the nozzle's suction action through the I/O card in robot controller. It sorts workpieces of different shapes into different locations, and then comes back to the initial position (Fig. 7).

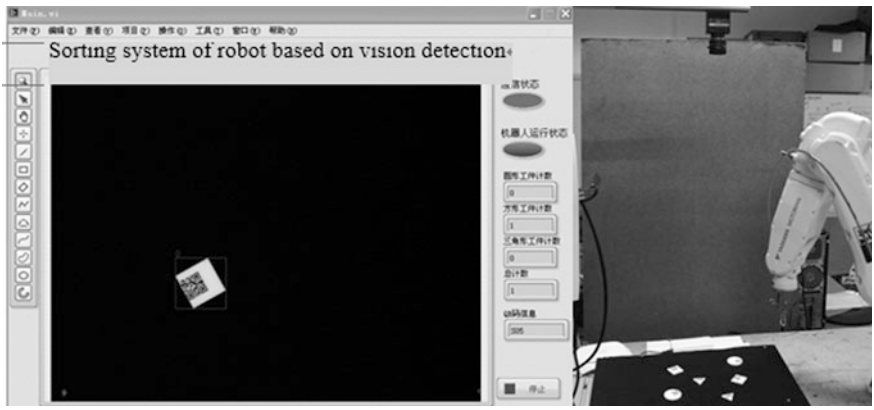


Fig. 7 Operation effect

5 Conclusion

This article describes a visual-inspection-based robot sorting system. The results of practice have demonstrated that this system is capable of realizing unmanned sorting operation with high efficiency and safety to replace heavy manual labor and improve automation level and operation quality. The construction and operation of the visual-inspection-based robot sorting system have accumulated experience for further research and development, and provided reference for implementation of robot sorting systems in production fields.

Acknowledgements This work is sponsored by key disciplines of mechanical engineering of Shanghai Polytechnic University (XXKZD1603), Graduate project fund (EGD16YJ025).

References

1. Jiao EZ, Du R (2010) Realization of industrial robot sorting technology. Modul Mach Tool Autom Manuf Techn 02:84–87
2. Liu ZY, Zhao B, Zou FS (2012) Application of machine vision technology in workpiece sorting. Comput App Softw 11:87–91

The Challenges and Promises of Big Data —An Engineering Perspective

Yi Wang

Abstract The term "big data" have been used for the analysis of data with high volume and veracity. Typically the size of the data is so large that commercial processing software, such as Excel and SPSS is inadequate to deal with them. This paper evaluates and examines the proper means of applications in which big data can be successfully deployed. Special attention is given to product design, which is one of the most recent of the big data approaches. Current trends and recent developments in big data analytics research are also discussed. The paper concludes with a summary of some of the key research issues in big data analyses.

Keywords Big data · Interoperability · Data storage · Data management
Sense making

1 Introduction

The term "big data" has been in use since the 1990s [1]. Big data is defined as large data sets beyond the handling-capacity of commercial software and process within a scalable time frame [2, 3] Big data "size" is relative to enterprise capacity, a small enterprise might find difficult to handle terabytes of data as google/SQL can deal with petabytes [4, 5]. Gartner [6, 7] identified that Big data can be transformed into information and valuable knowledge [8, 9].

Y. Wang (✉)
School of Business, Plymouth University, Plymouth, UK
e-mail: Yi.wang@plymouth.ac.edu

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_63

2 Criticism of Big Data Paradigm

Critiques of the big data paradigm have been developed from many fronts, for example on the implications of the concepts [10], on the current development [11] and the critical data studies [12]. Often, the skills are ignored, for employees who are applying the data [12, 13].

Graham [14] has criticised the focus big data usage are mostly on the economic, contexts. Most algorithms presumed that the future is similar to the past [15, 16]. Privacy is one of the major concerns around the analysis of Big data [16, 17]. The use of Big Data should be monitored at all levels both domestically and internationally [18, 19], to protect individual security from large enterprise or government's misuse the information. Several researches argued the ethical and governmental handling the big data [20, 21]. Recent developments in Business intelligence domain, demonstrated that through the analysis of smaller data or part of a data sets, would generate sufficient results [22–24].

Big data has been criticized as a buzzword [24, 25], this comes mostly from consultants and other practitioners [26, 27]. Results from Big data analytics are dominated by the quality of the analysis model [28] (Fig. 1).

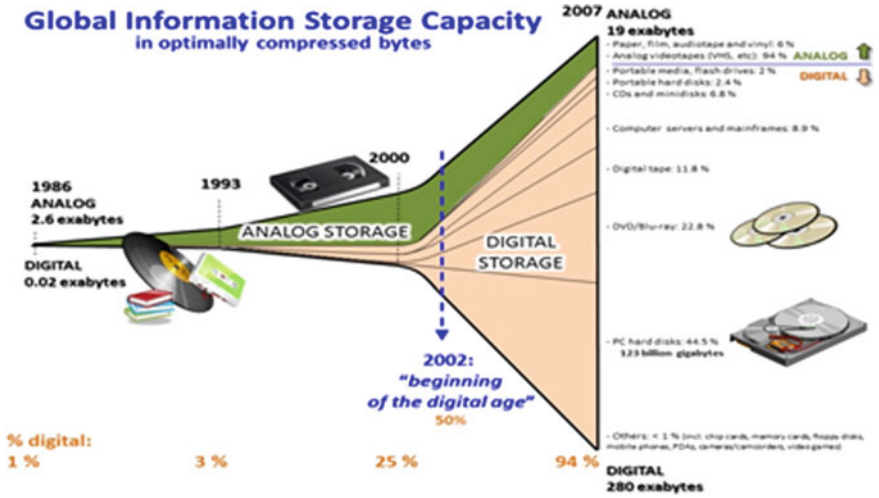


Fig. 1 Big data paradigm [10]

3 New Big Data Paradigm

This part the Big data triad have been explicitly identified as an object of study. Big data triad typically involve three distinct functions: Storage, Analysis and inter-operability. There may be other functions related to Big data which can be formed differently. However, our specific interest here is in the Storage, Analysis and interoperability triad, as shown in Fig. 2.

Within the domain of Big data triad focuses on shifting relationship structures, and in particular on the position of the Analysis vis-à-vis that of the Storage and Interoperability [29]. In the triad, the 'bridge' position of the Analysis between storage and Interoperability, providing information and monitoring benefits, may decay as the Storage become incompatible with the Interoperability, leading to erosion of those benefits. By investing in communication with Storage and Interoperability will yield information on how the storage performs and enabling the Analysis to better control performance.

Many processes link Storages and Analysis directly with Interoperability. The Storage's capabilities are contingent on access to the field Big data sites and installed base information, which require collaboration with Interoperability. Triadic cooperation between Storage, Analysis and Interoperability improves the value to the Interoperability.

The suitability of different forms of consistent collaboration for the Analysis-storage and Analysis-Interoperability can be investigated the conditions for Big data delivery quality control in different phases (e.g. Data design, Data/knowledge delivery). Various aspects of Big data quality and the different roles that Big data triad actors have in managing quality. As shown in Fig. 3 the relation between data and storage [30].

The triad is implicit; the research on triads has identified on variety of triad structures. However, most Big data triad studies have focused on the process of relationship formation and dissolution in triad. The big data triads tends to down-play both routine action and tacit forms of knowing, whereas a considerable volume of work on organisations [31]. The use of triad suggests the need for more focus on how the big data have changed the ways in which actors within organizations use

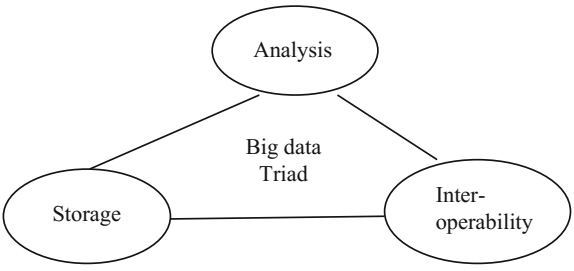


Fig. 2 Big data triad

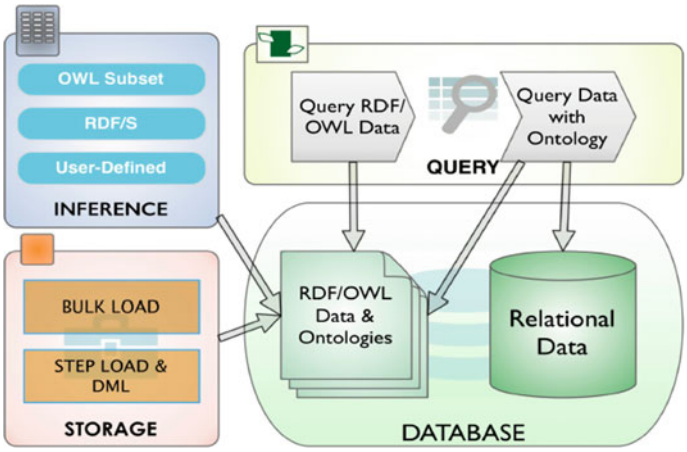


Fig. 3 Big data storage [30]

information. This is challenging, because it falls at the intersection of a number of disciplinary domains. It also suggests that the use of triad as a framework over time, as opposed to looking at how algorithms are implemented.

The Big data triad is central to the activities of multinational organizations and to the way in which those who work in them use the information. The big data triad should from a system which reveal over a long time line the processes and mechanisms that explain how the ICT for development trajectory emerged from the interplay of interactions at different levels of players in global, national, public, private, third sector, and civil society institutions and organizations.

The interoperability enabled separation of the parts from the whole, the analysis of time and temporality in processes of transformation, and the role of network structures and dynamics in explaining information/knowledge-related change. A key challenge here is a conceptualization of the coordinating process based on a critical realist perspective. Using a logic employing deductive, inductive, and retroductive elements, that help to explain the coordinating process and outcomes observed in the two efforts.

References

1. Lohr S (2013) The origins of 'big data': an etymological detective story. New York Times. http://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/. Retrieved 28 Aug 2017
2. Snijders C, Matzat U, Reips U-D (2012) 'Big data': big gaps of knowledge in the field of internet. Int J Internet Sci 7:1–5
3. Dedić N, Stanier C (2017) Towards differentiating business intelligence, big data, data analytics and knowledge discovery (285). Springer International Publishing, Berlin, Heidelberg

4. Everts Sarah (2016) Information overload. Distillations 2(2):26–33
5. Ibrahim TH, Abaker Y, Ibrar BA, Nor MS, Gani A, Ullah Khan S (2015) "Big data" on cloud computing. Inform Syst 47:98–115
6. Laney D (2011) 3D data management: controlling data volume, velocity and variety (PDF). http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. Retrieved 6 Feb 2017
7. Beyer M (2011) Solving 'big data' challenge involves more than just managing volumes of data. http://www.gartner.com/it/page.jsp?id=1731916. Retrieved 13 July 2017
8. De Mauro A, Greco M, Grimaldi M (2016) "A formal definition of big data based on its essential features". Libr Rev 65:122–135
9. Grimes S (2011) Big data: avoid 'Wanna V' confusion. InformationWeek. http://www.informationweek.com/big-data/big-data-analytics/big-data-avoid-wanna-v-confusion/d/d-id/1111077? Retrieved 5 Jan 2017
10. Hilbert M, López P (2011) The world's technological capacity to store, communicate, and compute information. Science 332(6025):60–65
11. Kimble C, Milolidakis G (2015) Big data and business intelligence: debunking the myths. Glob Bus Organ Excellence 35(1):23–34
12. Anderson C (2008) The end of theory: the data deluge makes the scientific method obsolete. WIRED. https://www.wired.com/science/discoveries/magazine/16-07/pb_theory. Retrieved 5 Jan 2017
13. Graham M (2012) Big data and the end of theory?. The Guardian. London. https://www.theguardian.com/news/datablog/2012/mar/09/big-data-theory. Retrieved 5 Jan 2017
14. Shah S, Horne A, Capellá J (2012) Good data won't guarantee good decisions. Harvard Bus Rev 35(1):23–34
15. Hilbert M (2014) Big data requires big visions for big change, London: organized TED talks. https://www.youtube.com/watch?v=UXef6yfJZAI. Retrieved 5 Jan 2017
16. Rauch J (2002) Seeing around corners. The Atlantic. https://www.theatlantic.com/magazine/archive/2002/04/seeing-around-corners/302471. Retrieved 5 Mar 2017
17. Epstein JM, Axtell RL (1996) Growing artificial societies: social science from the bottom up. A Bradford Book, UK
18. Delort P (2012) Big data in biosciences, big data Paris. http://www.bigdataparis.com/documents/Pierre-Delort-INSERM.pdf#page=5. Retrieved 5 Jan 2017
19. Hawkins RD, Hon* GC, Ren B (2010) Next-generation genomics: an integrative approach. Nat Rev 11
20. Tambe SS (2015) "Big data in biosciences", insights in biology—2025. CSIR-National Chemical Laboratory, Pune, India, pp 25–28
21. Ohm P (2012) Don't build a database of ruin. Harvard Bus Rev. http://blogs.hbr.org/cs/2012/08/dont_build_a_database_of_ruin.html
22. Wares F (2010) Failure to launch: from big data to big decisions. http://www.fortewares.com/Administrator/userfiles/Banner/forte-wares–pro-active-reporting_EN.pdf
23. Pelt M (2015) "Big Data" is an over used buzzword and this Twitter bot proves it. siliconangle.com. SiliconANGLE. http://siliconangle.com/blog/2015/10/26/big-data-is-an-over-used-buzzword-and-this-twitter-bot-proves-it/
24. Gregory P (2014) Interview: Michael Berthold, KNIME Founder, on research, creativity, big data, and privacy, part 2. KDnuggets. http://www.kdnuggets.com/2014/08/interview-michael-berthold-knime-research-big-data-privacy-part2.html
25. Harford T (2014) Big data: are we making a big mistake? Financial Times. http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html
26. Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8):e124. PMC 1182327
27. Lohr S, Singer N (2016) How data failed us in calling an election. The New York Times. ISSN 0362-4331. https://www.nytimes.com/2016/11/10/technology/the-data-said-clinton-would-win-why-you-shouldnt-have-believed-it.html

28. Markman J (2016) Big data and the 2016 election. Forbes. http://www.forbes.com/sites/jonmarkman/2016/08/08/big-data-and-the-2016-election/#4802f20846d7
29. Calvanese D, Cogrel B, Komla-Ebri S, Kontchakov R, Lanti D, Rezk M, Rodriguez-Muro M, Xiao G (2017) Ontop: answering SPARQL queries over relational databases. Semant Web J 8:471–487
30. Oracle semantic technologies developer's guide 11 g release 2. Available online: http://docs.oracle.com/cd/E11882_01/appdev.112/e25609/title.htm. Accessed on 1 Dec 2016
31. Mutch A (2010) Technology, organization and structure—a morphogenetic approach. Organ Sci 21(2):507–520

A Modified Teaching and Learning Based Optimization Algorithm and Application in Deep Neural Networks Optimization for Electro-Discharge Machining

Chen Wang, Baorui Li, Yi Wang, Kesheng Wang and Shenghuai Wang

Abstract In order to improve the output precision of depth neural networks, an improved teaching and learning optimization algorithm is proposed to optimize the weights and thresholds of depth neural networks. The algorithm is improved according to the teaching and learning phases of the basic teaching and learning algorithms. The performance of the algorithm is tested by electro-discharge machining (EDM) experiments. The results show that the algorithm has the advantages of fast convergence and high solution accuracy.

Keywords Teaching and learning · Deep neural networks · Electro-discharge machining

1 Introduction

In recent years, deep neural networks (DNN) has achieved good results in various fields [1]. The shallow layer neural network is easy to appear the following problems: 1, over (under) fitting phenomenon; 2, model optimization objective function is highly non convex, affect the stability and convergence of the network model; 3, back-propagation algorithm (BP) using the gradient descent method is by updating the parameters will lead to gradient diffusion phenomenon.

DNN provides a more reliable and simple method for solving large-scale complex problems [2]. DNN is divided into input layer, hidden layer and output layer. The input layer node is used to receive a set of variables, and the output layer

C. Wang (✉) · B. Li · K. Wang
Shanghai Key Laboratory of Intelligent Manufacturing and Robotics, Shanghai University, Shanghai, China
e-mail: wangc_jx@huat.edu.cn

C. Wang · S. Wang
Hubei Automotive Industries Institute, Shiyan, China

Y. Wang
School of Business, Plymouth University, Plymouth, UK

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_64

outputs the corresponding objective value. DNN mainly regulates the connection weights between the input layer, the hidden layer and the output layer, so that the output error of the network and the actual output of the problem are as small as possible [3].

Therefore, how to adjust the network connection weight becomes the research hotspot of DNN structure. At present, the optimization of neural networks is mainly gradient optimization algorithm, but because the gradient descent method requires higher initial values, the multi-peak optimization problem is easy to fall into the local optimum and lose the global optimal solution. In recent years, many intelligent optimization algorithms have been used in the optimization design of DNN, such as algorithm (genetic) and so on. This paper proposes an improved teaching and learning optimization algorithm (teaching and Learning based optimization, TLBO) to optimize the design of DNN.

Among known non-conventional machining methods that nowadays have a wide range of applications on the production floor, electro-discharge machining (EDM) is the most extensively used technique for precision machining and manufacturing of prototypes. EDM is an electro thermal process where the material removal mechanism is achieved by the erosive effects from repetitive electrical sparks generated between tool (the anode) and work material (the cathode) with constant electric field emerged in dielectric fluid. Figure 1 shows a simplified diagram of the working principle of EDM process. In this process, the feed motion of the tool is controlled by a servo-controller that maintains a spark gap in the range of 0.025–0.05 mm. The workpiece is placed in the dielectric fluid that circulates through a hole or holes in the tool electrode under pump pressure.

The process is also one of the expertise-demanding processes in the industry, and the mechanism of metal erosion during sparking is still debatable due to the complex thermal conduction behaviors in the machining vicinity. This may explain the reason why it is hard to establish models that accurately correlate the process variables and performances. Existing practices have difficulties to identify as which parameters need to be changed or how they should be modified for best process

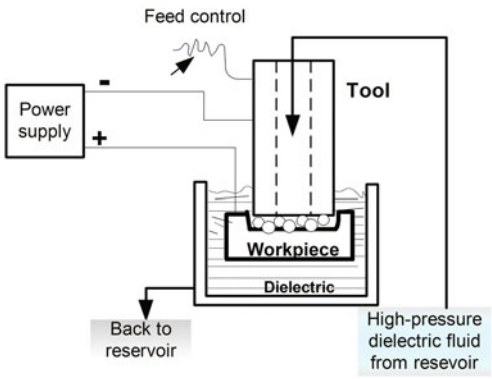


Fig. 1 The EDM machine

performance. In addition, when new and advanced materials appear in the field, it has not been possible to use existing models.

A EDM modeling method was used back propagation neural network is proposed using back propagation neural network by Mandal et al. [4]. Ming et al. proposed soft computing models and intelligent optimization system in electro-discharge machining of sic/al composites [5]. Xiuli et al. was used bp neural network l-m algorithm to process parameter optimization of ti-6al-4v electric discharge machining [6].

Modeling EDM for graphite tool material was mainly researched by this paper. The modeling of this process is a special problem, and it is impossible to define the objective function in a manufacturing environment with a smooth continuous mathematical formulation. By combining the improved teaching and learning algorithm with the depth neural network, a hybrid method for solving modeling and optimization problems is developed. The input–output model is solved using the experimental data. To achieve the target of the task, a two-phase hybrid system of Mtlbo algorithm and deep neural network has been implemented. These two phases can be categorized as modeling phase and optimization.

The remaining of this paper is organized as follows. Section 2 presents the TLBO, MTLBO, DNN and DNN optimization based on MTLBO. Section 3 includes the experimental studies for demonstrating the efficiency of the proposed method in addition to some discussions,and the conclusions are presented in Sect. 4.

2 Teaching and Learning Based Optimization and Deep Neural Networks

2.1 Teaching and Learning Based Optimization

Teaching and learning based optimization is an important process where every individual tries to learn something from other individuals to improve themselves [7–11]. The original TLBO is simulated with the class as a unit of learning, in the teaching stage, students through the teacher's teaching and learning, which makes their knowledge to the teacher together, at the same time, to strengthen the students between the absorption of knowledge through learning from each other. Among them, teachers and students are equivalent to individuals in the group of intelligent optimization algorithms, and teachers are one of the best individuals, each student is equivalent to a decision variable.

(1) Algorithm initialization

Class: the equivalent of the population, for all the students of the collection, recorded as $X(x_1, x_2 \ldots x_n)$, n is student number. Student: Equivalent to the individual, as a class member, recorded as $X_i = (xi_1, xi_2 \ldots xi_d)$, $x_{i,j}(i = 1, 2, 3 \ldots n;$

j = 1, 2, 3… d), X_{ij} represents the jth course of student X_i. Teacher: The best individual, the best of the students, recorded as $X_{teacher}$. Search for area boundaries: $X_{i,j} = (x_{i,1}, x_{i,2} \ldots x_{i,d}) \in [L, U]$, $L = (L_1, L_2, \ldots L_d)$, L is the lower bound of the search region, $U = (U_1, U_2, \ldots U_d)$, U is the upper bound of the search area, and d is the number of courses by the student has learned, that is, the dimension corresponding to the optimization problem. For the multi-objective optimization problems, Minimize $F(x) = (f_1(x), (f_2(x) \ldots (f_n(x))^T$.

Class for students is x_j (j = 1… PN), $x_{teacher} = \min f(x_j)(j = 1, \ldots, PN)$. NP for the number of students, and d for the number of subjects studied. Then, the population is initialized, the first i student's j is (Fig. 2):

$$x_{i,j}^0 = L_j + rand \times \left(U_j - U_l\right) \tag{1}$$

The teaching formula is as follows:

$$x_i^{new} = x_{old}^i + difference \tag{2}$$

$$difference = r_i \times \left(x_{teacher} - TF_i \times mean\right) \tag{3}$$

X_{old}^i and x_{new}^i that the value of the i students before and after learning respectively, mean = $\frac{1}{NP}\sum_{i=1}^{NP} X^i$ is the average of all student grades. TFi = round [1 + rand (0, 1)] is learning factor, r_i = rand (0, 1) is learning step.

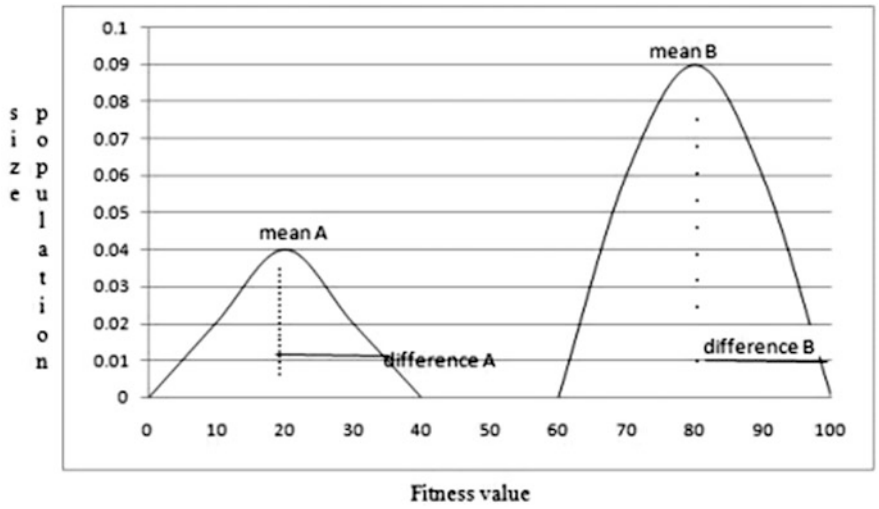


Fig. 2 Teaching method

(2) Update student

$$\text{if f}\left(x_{\text{new}}^{i}\right) > \text{f}\left(X_{\text{old}}^{i}\right)$$
$$x_{new}^{i} = X_{old}^{i} \tag{4}$$

(3) Learning phase

$$x_{new}^{i} = X_{old}^{i} + \text{r}_i \times \left(\text{X}_i - \text{X}_j\right) \quad \text{f}(\text{X}_\text{i}) < \text{f}\left(\text{X}_\text{j}\right)$$
$$\text{X}_{old}^{i} + \text{r}_\text{i} \times \left(\text{X}_\text{j} - \text{X}_\text{i}\right) \quad \text{otherwise} \tag{5}$$

(4) Termination condition

If the end condition is satisfied, the optimization is completed, otherwise go to 2 Continue.

2.2 Deep Neural Networks

As mentioned in the previous section, the hidden and visible variables are not mutually independent. To explore the dependencies between there variables, in 2006, Hinton constructed the DBNs by stacking a bank of RBMs (Restricted Boltzmann Machine). The deep belief network (DBN) is stacked by several layers of bounded Boltzmann machines (RBM), and the hidden layer on the upper RBM layer serves as the visible layer of the next RBMs [12].

The DBN model is stacked with several layers of RBM, and if there is tag data in the training set, then the last layer of RBM's visible layer contains both the hidden layer cell of the previous layer RBM and the tag layer cell. The top layer of RBM is visible assuming 500 neurons, classification of the training data is divided into 10 categories, then the top layer visible RBM has 510 dominant neurons, each of the training data, the corresponding label neurons are open to 1, while the others were close to 0. The DBN model structure is shown in Fig. 3.

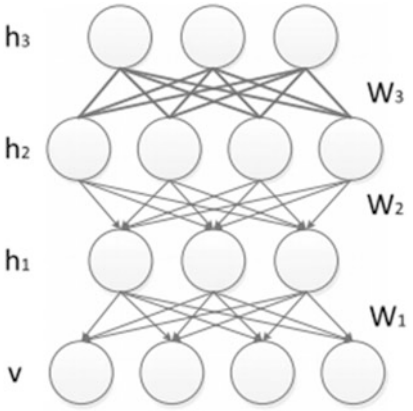


Fig. 3 The DBN model structure

2.3 A Modified Teaching and Learning Based Optimization Algorithm

In the traditional TLBO algorithm, the student grade is achieved through the teaching of a single teacher. However, if the majority of students in the class grade lower than the average grade of class, just by one teacher to teach students, it is difficult to improve the grade of all students. During this optimization process, this situation not only leads to the convergence of the algorithm is too slow and also easy to fall into the local optimal. Taking into account the above facts, the original TLBO algorithm needs to be improved.

(1) Grouping strategy

The grouping strategy is for grouping operations operations. The grouping operation is to divide all students in the class into IS students into MS groups, each with NS students, and each with a teacher in each group. The grouping criteria is to put the first place in the first group, the second student into the second group, the MS students into the MS group, and then MS + 1 students into the first group, and so on. The first 2MS students into the first MS group. Where KS group FKS is FKX = X (KS + MS (P − 1)), P = 1,…, NS, KS = 1…MX. And then calculate the fitness value of each student, the students according to the fitness value of the sort, while recording the global optimal student X_G and the group of the best students X_P. By means of introducing more than one teacher to teach students, each teacher can improve the results of its assigned group and if the level of the group reaches up to the assigned teacher, thereafter, a better teacher will be assigned to this group. During the process of optimization this strategy increases the diversity of the algorithm.

(2) Simulation annealing

SA is a meta-heuristic technique based on a thermodynamic process of the annealing of material. The SA algorithm is constructed based on the statistical mechanics, which was demonstrated by Metropolis et al. in 1953 through the concept of Boltzmann's probability distribution. It means if a system is maintained in a thermal equilibrium at temperature T, then the probability distribution p of its energy E can be achieved by [21]:

$$P(E) = e\frac{-\Delta E}{k_B T} \tag{6}$$

Where KB is a Boltzmann's constant. The difference in energy ΔE means the difference in cost function between the past and current iterations, which can be determined as follows:

$$\Delta E = f(x_n) - f(x_o) \tag{7}$$

For minimization problems, $\Delta E \leq 0$ means $f(x_n) \leq f(x_o)$, so the new design point is directly accepted. Otherwise, the Metropolis criterion will be enabled to decide whether to accept or reject Xo. For this case where $\Delta E > 0$, the acceptance is treated probabilistically according with the relation $P = \frac{1}{1+e^{(\Delta E/\max(T))}}$. It can be viewed the influence of temperature in the acceptance process. For the highest magnitudes of T, The acceptance probability to choose a worse state is likewise higher. This process will avoid trapping into local optima. As the temperature decrease, the SA algorithm accepts only states which minimize the FO cost. Therefore, the way that temperature decreases during the iteration of the algorithm is an important parameter, this parameter is named cooling schedule.

(3) Algorithm step

The Improved Teaching and learning based optimization (ITLBO) algorithm was first preliminary processing of NC cutting parameter, then optimize the function calculation, and then grouping into a plurality of candidate solutions group. During the teaching stage, according to (2) and (3) to calculate the value of teachers and update, Metropolis criterion was introduced to select poor solutions, the poor solution can also solve the enhancement algorithm diversity. Then according to (4) to carry out the learning stage to update the student solutions, the Metropolis criterion was introduced again, the final iteration of the final judgment, when the number of iterations to reach the set, the output of the final Pareto solution. The ITLBO algorithm as the Fig. 4.

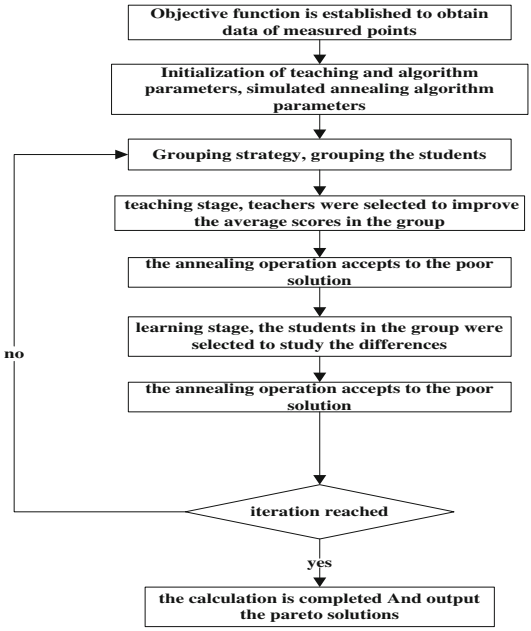


Fig. 4 ITLBO algorithm

2.4 DNN Optimization Based on Mtlbo

In this section, a two-phase hybrid system of Mtlbo algorithm and deep neural network has been implemented. In the Fig. 5, the first phase involves the establishment of the model using a multi-layer feedforward deep neural network architecture. Instead of the standard backpropagation error minimization approach, Mtlbo is implemented to find the optimum values of the weights that minimize the error between the measured and the evaluated (network output) performance parameters. In this phase of hybridization, the weight matrices between input and hidden nodes; and hidden and output nodes were coded as students. Then, the following relation was used to combine the inputs of the network at the nodes of the hidden layer [13–18].

3 Experimental Results

In this section, we compare the performance of MTLBO in real-world EDM problems with Original TLBO and Genetic algorithm optimization (GA) [19–21].

Many independent parameters influence the performance of EDM processes including Ontime Offtime peak current, voltage, compression gain, dielectric fluid, electrode material, material removal rate, surface roughness and tool wear. Table 1 shows the values of the input parameters for the test done on graphite electrode

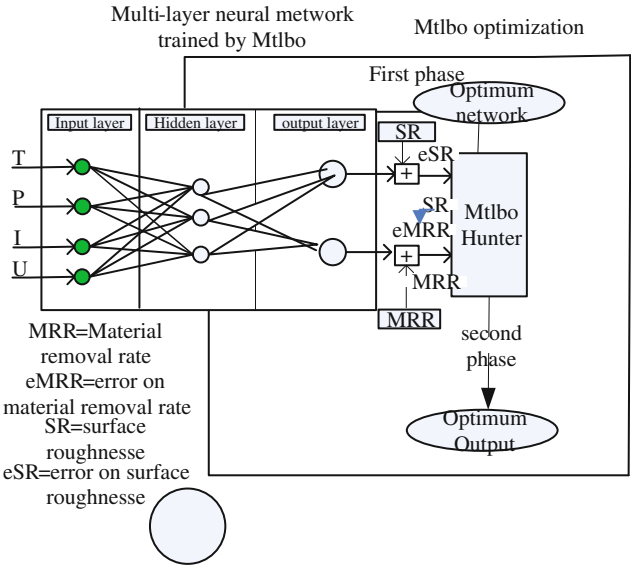


Fig. 5 DNN based on Mtlbo system for EDM process modeling and optimization

Table 1 Experimental dataset for graphite tool on nickel-base alloy

Input parameter				Performance parameter	
T	P	I	U	MRR	SR
56	10	21	100	380	7.87
42	10	21	100	682	6.37
32	10	21	100	1039	5.44
24	10	21	100	628	5.54
49	56	39	100	578	7.14
49	37	29	100	583	7.00
49	10	21	100	1345	6.87
49	37	21	100	1018	8.06

(size 2.9×9.8 mm^2) with nickel-base alloy workpiece using AGIE INNOVATION EDM machine.

For example, the material removal rate can be expressed as follows:

$$\text{Minimize} \quad \left(E_k = \sqrt{\sum_{k=1}^{21} (Y_k - Q_k)^2} \right)_{MRR} \tag{8}$$

where, E_k is the mean squared error between the output (Y_k) and the target performance (Q_k) of node k. Further, to optimize the error in the deep neural network

Table 2 GA, TLBO and MTLBO parameter definitions

Algorithm	P	I	CP	MP	J	T_0	T_{end}
GA	50	30	0.6	0.1	–	–	–
Tlbo	50	30	–	–	–	–	–
Mtlbo	50	30	–	–	5	1000	0

Table 3 The test results of three algorithms on the EDM experiment

Algorithm	Test sample error	Training sample error	Prediction accuracy/%
GA	0.00512	0.01425	97
Tlbo	0.02418	0.05372	94
Mtlbo	0.00015	0.00028	99

structure and define the correlation, We use GA, TLBO and MTLBO algorithms to optimize the neural network, it was used with the following the mainly parameter definitions and shown in Table 2. (population:p, Iteration number:I,Cross probability:cp,Mutation probability:mp,Class group:J,initial temperature:T_0,End temperature:T_{end}).

Table 3 shows that the neural network model of this algorithm is less error. Compared with other algorithms, the algorithm model is accurate.

4 Conclusions

In this paper, an improved teaching and learning algorithm based on DNN is proposed in this paper. The algorithm is used to optimize weights and thresholds of deep neural networks, so that the error of network output and expected output is minimum. Experiments were carried out to verify the EDM process, and it was showed that the proposed algorithm has obvious advantages over the genetic algorithm and the basic TLBO algorithm.

Acknowledgements This research work was supported by National natural science foundation of china (No.51475150) and Key Laboratory of Automotive Power Train and Electronics (Hubei University of Automotive Technology, No.ZDK1201703).

References

1. Sze V, Chen YH, Yang TJ, Emer J (2017) Efficient processing of deep neural networks: a tutorial and survey
2. Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE (2016) A survey of deep neural network architectures and their applications ☆. Neurocomputing 234:11–26
3. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85

4. Mandal D, Pal SK, Saha P (2007) Modeling of electrical discharge machining process using back propagation neural network and multi-objective optimization using non-dominating sorting genetic algorithm-ii. J Mater Process Tech 186(1–3):154–162
5. Ming W, Ma J, Zhang Z, Huang H, Shen D, Zhang G et al (2016) Soft computing models and intelligent optimization system in electro-discharge machining of sic/al composites. Int J Adv Manuf Technol:1–17
6. Xiuli MA, Teng K (2016) Process parameter optimization of ti-6al-4v electric discharge machining based on bp neural network l-m algorithm. Mater Rev
7. Anitha J, Dasa R, Pradhan MK (2016) Multi-objective optimization of electrical discharge machining processes using artificial neural network. Jordan J Mech Ind Eng 10(1)
8. Velázquez-Iturbide JA, Debdi O, Paredes-Velasco M (2015) A review of teaching and learning through practice of optimization algorithms
9. Rao RV, Patel V (2012) An elitist teaching-learning-based optimization algorithm for solving complex constrained optimization problems. Int J Ind Eng Comput 3(4):535–560
10. Rao RV, Savsani VJ, Vakharia DP (2011) Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems. Comput Aided Des 43 (3):303–315
11. Rao RV, Kalyankar VD, Waghmare G (2014) Parameters optimization of selected casting processes using teaching–learning-based optimization algorithm. Appl Math Model 38 (23):5592–5608
12. Liu J, Gong M, Miao Q, Wang X, Li H (2017) Structure learning for deep neural networks based on multiobjective optimization. IEEE Trans Neural Netw Learn Syst PP(99):1–14
13. Warghat ST, Deshmukh TR (2017) Parameter optimization of milling operation using teaching-learning-based optimization and artificial neural network. MAYFEB J Mech Eng 1
14. Kankal M, Uzlu E (2016) Neural network approach with teaching–learning-based optimization for modeling and forecasting long-term electric energy demand in Turkey. Neural Comput Appl:1–11
15. Shouheng T (2013) A modified teaching-learning-based optimization algorithm and application in neural networks. Caai Trans Intell Syst 8(4):327–332
16. Theodoridis S (2015) Chapter 18—neural networks and deep learning. Mach Lear:875–936
17. Maity K, Mishra H (2016) Ann modelling and elitist teaching learning approach for multi-objective optimization of μ-EDM. J Intell Manuf:1–18
18. Wang H, Shao F, Li H, Wang X, Li X, Department ME et al (2016) Optimization of electrical parameter in electrical discharge machining of engineering ceramics based on fuzzy neural network. Mod Manufact Eng
19. Napitupulu R, Wahyudi A, Soepangkat BOP (2015) Performance characteristics optimization of electrical discharge machining process using back propagation neural network and genetic algorithm 25(3)
20. Rao RV (2016) Multiobjective optimization of machining processes using NSTLBO algorithm. In: Teaching learning based optimization algorithm. Springer International Publishing
21. Wang C, Wang Y, Wang K, Dong Y, Yang Y (2017) An improved hybrid algorithm based on biogeography/complex and metropolis for many-objective optimization. Math Probl Engineering 2017 (2017-3-30):1–14

A Study on a Novel Application of Eye Tracking Technology in Product Customization

Baorui Li, Yi Wang, Kesheng Wang, Jinghui Yang and Lilan Liu

Abstract In the context of industry 4.0, personalized customization of products will be an important part of intelligent manufacturing. In a market competition, improving customer satisfaction is important to enhance the competitiveness of enterprises. Acquired demand information scientifically is significant to improve customer satisfaction. In this paper, we choose Tobii eye-tracking glasses as a research tool, the key chain and the customizable pictures as research objects. We selected 30 college students as subjects to repeat the experiment. The experiment collect the eye-tracking data of the subjects in and extract the fixation time of the subjects as a reference feature. We combine the relevant knowledge of statistical analysis to study the possibility of the rapid extraction of customer demand information by the eye tracking technology.

Keywords Industry 4.0 · Demand information · Personalized customization
Eye tracking · Fixation time

B. Li · L. Liu
Shanghai Key Laboratory of Intelligent Manufacturing and Robotics,
Shanghai University, Shanghai 200072, People's Republic of China

Y. Wang (✉)
School of Business, Plymouth University, Plymouth, UK
e-mail: yi.wang@plymouth.ac.uk

K. Wang
Knowledge Discovery Laboratory, Norwegian University of Science and Technology,
Trondheim, Norway

J. Yang
Department of Industry Engineering, Shanghai Polytechnic University, Shanghai, China

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_65

1 Introduction

In the vision of industry 4.0, Smart factories are no longer like a set of classic assembly lines or a workshop. A true smart factory is a full process ecosystem that includes design, procurement, manufacturing, marketing and services, and is a system that meets consumer's demand [1]. With the full integration of the internet plus and manufacturing, the manufacturing mode is constantly transforming and upgrading. Intelligent manufacturing mode provides material basis and technical support for personalized customization services [2]. Personalized business models will also become more popular, it is the core content of intelligent manufacturing [3]. How to carry out personalized customization service has become the focus of business attention [4]. An enterprise can pre-empty the market only by take the lead in launching a personalized custom service.

The enterprise's ability of survival and development with customer satisfaction have a positive proportion in a certain range [5]. As shown in Fig. 1. If an enterprise wants to maintain its strong development potential and in an invincible position in the competition. Apart from providing reliable and inexpensive products, enterprises are more important to meet customers' needs and respond quickly [5]. Personalized customization respects the personalized needs of each customer, and it maximizes the user demand by carrying out differentiated services to each user [6].

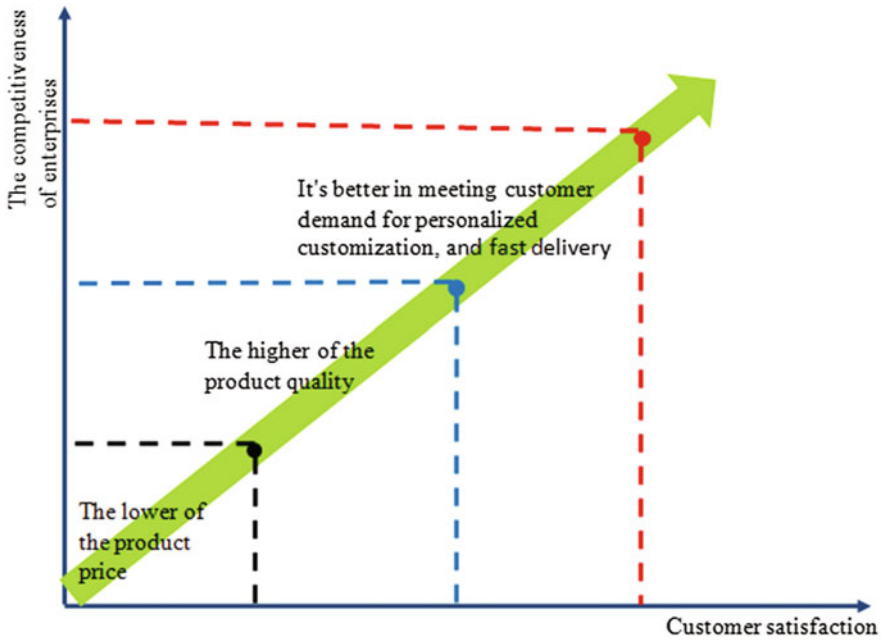


Fig. 1 The relationship between enterprise competitiveness and customer satisfaction

2 Literature Review

2.1 Personalized Customization

A personalized customization service is customer-centric, it can quickly and accurately acquire the personalized requirements of customers and push the appropriate customization options according to their own characteristics [5]. Finally we selected the product satisfied by the customer through the simple and smooth interactive customization operation. It is important to realize and satisfy the design function of new products [6]. The enterprise should takes advantage of the technology and understands the market to makes the overall design plan, communicate with the customer timely in the design process, understands the customer's demand and tries to improve the customer satisfaction [3–5]. For a relatively mature product, the product quality stable, the prices are transparent and the features of the product have tend to be perfect, at this time on the operating strategy of product, how to improve customer satisfaction play a leading role [4]. How to improve customer satisfaction quickly, first of all, enterprises should understand customer requirements, adjust relevant products in time according to customized changes, and better meet customers' demands [7].

Customer demand is a custom requirement expressed by customers, which is the direct driving force of product development and the basis and source of product configuration and variation design [8]. The extraction, expression and transformation of customer demand are key to product customization [8]. The development of new products includes product definition, product design and process design [9]. The customer requirements are then converted into functional requirements and further translate into design parameters and the final process variables [7]. In the process of product customization, the customer information needs to be translated into requirements details and further transformed into product configuration solutions. The effective acquisition of customer demand is the primary problem that needs to be solved in the personalized customization [9]. The enterprise provides the corresponding products and services in a targeted way to improve customer satisfaction. Traditional methods of acquiring customer demand include market investigation, enterprise information analysis and online acquisition [9]. Intelligent and efficient access to the customer demand information is an important field of study, in the complicated market competition environment will be conducive to enterprise fast response, to a certain extent, realize customer personalization, and finally enhance the competitiveness of enterprises [8]. It will be conducive to the rapid response of enterprises in the complex market competition environment, to achieve personalized customization of customers in a certain extent, and ultimately improve the competitiveness of enterprises [7].

2.2 Eye Tracking Technology

Human access to information relies heavily on vision, and about 80–90% of the external information is obtained through human eyes [10]. Eye movement can represent the cognitive activity of a person, such as the reasoning and solving process of a user for specific tasks [11, 12]. In recent years, eye-tracking technology based on the digital eye movement video has become increasingly mature [13]. It uses the infrared camera device of high frequency sampling to capture the eyeball image in real time and further uses image processing to obtain the eye-tracking data [13].

There are three main forms of eye movement: Fixations, Saccades and Smooth pursuit [14–16].

(1) Fixations

The fixation is a stop on the observed target, these stops generally last at least 100–200 ms. In the process of fixations, the eye is not absolutely stationary and the eyeball constantly makes slight jitter to see the object, the magnitude of the jitter is generally less than 1°. The vast majority of information is acquired and processed only when it is being watched [14, 15].

(2) Saccades

The saccades is a rapid jump between the fixation points, which is a combined eye movement (it's the simultaneous movement of the two eyes). The perspective is 1°–40°, duration is 30–120 ms and the maximum speed is 400°–600°/s. During the period of this eye movement, the image moved too fast and the visual threshold of the image is raised when the eye beats in the retina, So almost no information is obtained in this process [14–16].

(3) Smooth pursuit

Eye gaze smooth smoothly moving at 1°–30°/s follow the target movement. This slow combined eye movement is commonly known as smooth tracking. The premise of smooth pursuit is that it must have a slow moving target [14, 16].

Human vision is characterized by directness, naturalness and bidirectional nature. Other senses don't have these advantages, so the eye-tracking technology plays an important role in cognitive science and psychology research [11, 17]. Along with the technology is gradually mature, eye tracking not only improves the sampling accuracy and frequency of collected eye tracking data, but also does not interfere with the normal operation of users [17]. The application eye tracking technology is more and more extensive, it has been widely used in usability assessment, human-computer interaction and other fields.

3 Materials and Methods

3.1 Stimuli

In the experiment, we selected six different colors key chains and six different cartoon pictures as the experiment objects. The experiment objects was divided into two groups: A group is six key chains of different colors. As shown in Fig. 2a, We use the set K to represent the key chains, so K = (k_1, k_2, k_3, k_4, k_5, k_6), where k_1 stands for blue key chain, k_2 for yellow, k_3 for red, k_4 for green, k_5 for green, k_5 for black, and k_6 for orange. We use the set P to represent the cartoon pictures, so P = (p_1, p_2, p_3, p_4, p_5, p_6). Details are shown in Fig. 2b. In the process of experiment, We use eye-tracking glasses to collect subject's eye-tracking data on the characteristics of each object, and chose the fixation time to measure subjects' interest in stimulating objects.

3.2 Subjects

We chose 30 college students as subjects. They are 15 boys and 15 girls, their age range is 18–24 years old. All subjects has normal color vision. The subjects were represented by 't', and the subjects set was represented by 'T', so T = (1, 2, 3 … 30).

3.3 Methods

The experiment selected Tobii eye tracking glasses as an eye-tracking data acquisition tool. The experimental process is shown in Fig. 3. The subjects looked at two groups for 2.5 s, eye tracking glasses record the eye movement tracking data of

Fig. 2 Stimuli

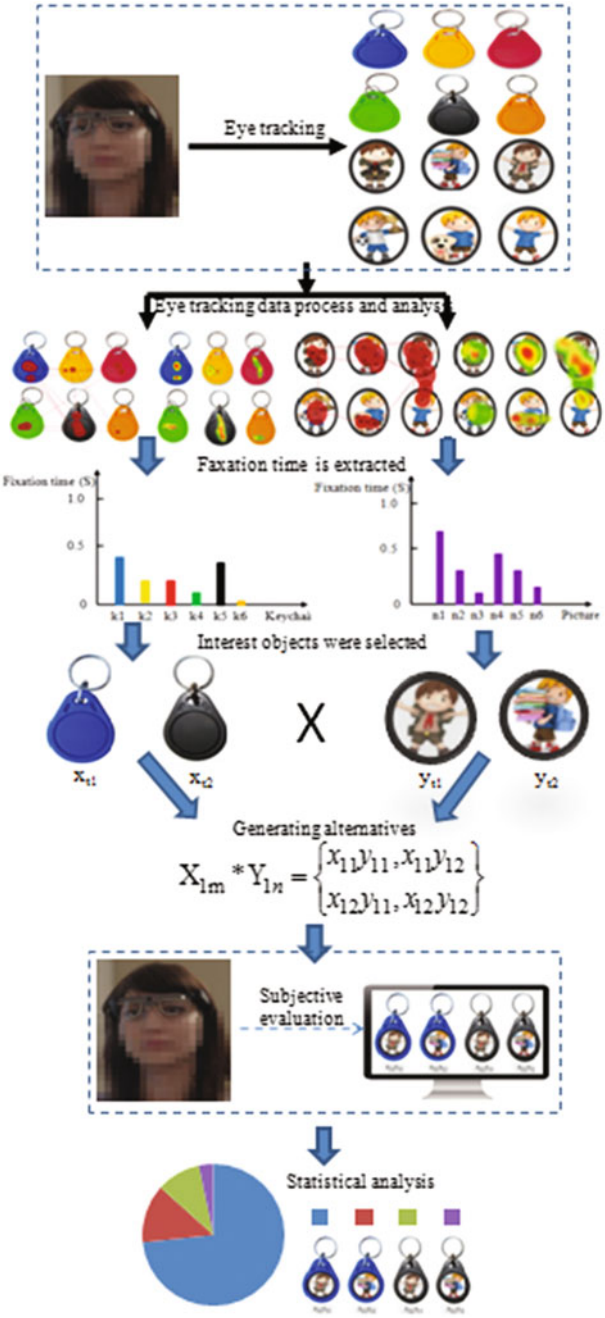


Fig. 3 The framework of our experiment processing

subjects in this process. We analysis and process the collected eye tracking data. Stimulus objects' fixation map and heat map are generated, these two types of maps directly show the subjects' interest in the stimulus objects. We selected the subjects' fixation time as a reference feature to quantify the interest of subjects in the stimulus objects. In each experiment, the participants' fixation time of each stimulus object was calculated, and the stimulus object was ordered by the fixation time. We select the two stimuli with the longest duration of fixation time, use x_{t1} to represent the key chain of first interest of the subject 't', and x_{t2} represents the key chain of second interest of the subject 't'. In the same way, we use y_{t1} and y_{t2} respectively represents the first interest and the second interested custom picture of the subject.

We cross-group the selected key chain and custom picture to generate recommendation scheme. Each subject will be given a set of eye-tracking data, Similarly, according to the eye movement data of e ach group, the corresponding recommendation scheme will be obtained. We make statistical analysis of the results and finally reach the conclusion.

4 Results and Discussion

4.1 Eye Tracking Data Analysis

During the experiment, eye-tracking glasses were used to collect the eye movement data of when subjects watching the stimulus objects. We analysis and process eye tracking data, the Fig. 4a shows the fixation map of the subject 't' on the key chain and he Fig. 4b shows the fixation heat map of the subject 't' on the key chain.

We extracted the subject's fixation time on different color key chain from eye tracking data, formed the Table 1. We use the data from Table 1 to generate the columnar figure, shows in Fig. 5. We can directly see the subject's fixation on the key chain of different colors from the fixation time columnar figure. The subjects' fixation time on different color key chains reflected the interest of the subjects in different color key chains.

Fig. 4 Fixation map and heat map of the key chain

Table 1 Fixation time of the keychain

	k_1	k_2	k_3	k_4	k_5	k_6
Fixation time	0.422	0.248	0.256	0.163	0.397	0.013

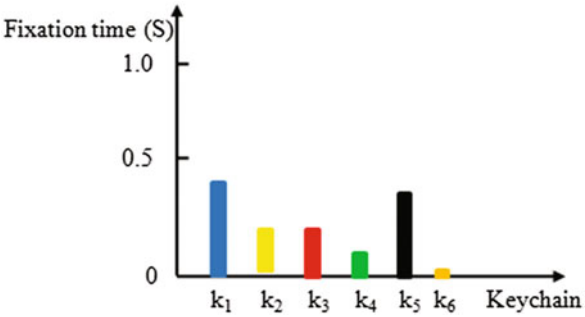


Fig. 5 The fixation time columnar figure

We used the same method to analyze the fixation time of each custom picture as the eye movement data of the subjects in the custom picture. The Fig. 6a shows subjects' fixation map of different customized pictures and the Fig. 6b shows the heat map of the subject 't' on different customized pictures.

Similarly, we extracted the fixation time of subjects from eye movement data to different custom pictures. Table 2 shows the extracted fixation time. At the same time, we use the data from Table 2 to generate the columnar figure, shows in Fig. 7.

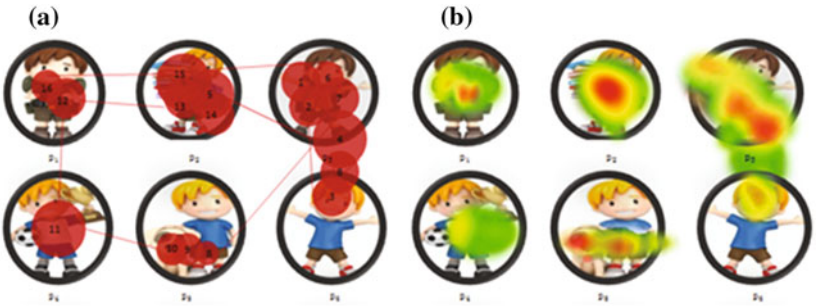


Fig. 6 Fixation map and heat map of custom pictures

Table 2 Fixation time of the custom pictures

	p_1	p_2	p_3	p_4	p_5	p_6
Fixation time	0.251	0.482	0.693	0.316	0.321	0.223

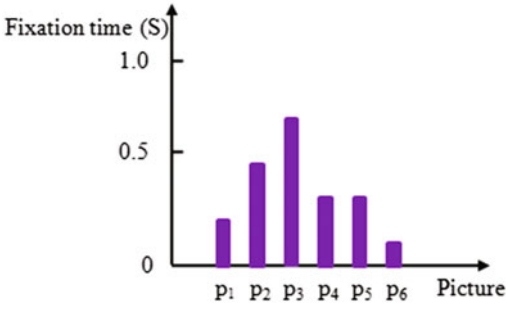


Fig. 7 The fixation time columnar figure

We can directly see the subject's fixation on the custom pictures from the fixation time columnar figure. The subjects' fixation time on different custom pictures reflected the interest of the subjects in different custom pictures.

4.2 Recommendation Scheme

According to the analysis results of eye movement data, the time of fixation was used as the measurement criterion, we use x_{t1} to represent the key chain of first interest of the subject 't', and x_{t2} represents the key chain of second interest of the subject 't'. In the same way, we use y_{t1} and y_{t2} respectively represents the first interest and the second interested custom picture of the subject. We cross-group the selected key chain and custom pictures to generate the recommendation scheme (look at the Eq. 1). Each subject will be given a set of eye-tracking data, Similarly, according to the eye movement data of each group, the corresponding recommendation scheme will be obtained. Recommendation scheme is shown in Fig. 8.

$$X_{1m} * Y_{1n} = \left\{ \begin{matrix} x_{11}y_{11}, x_{11}y_{12} \\ x_{12}y_{11}, x_{12}y_{12} \end{matrix} \right\} \tag{1}$$

Fig. 8 Recommendation scheme

Table 3 Statistical table

	$x_{11}y_{11}$	$x_{11}y_{12}$	$x_{12}y_{11}$	$x_{12}y_{12}$
Amount	22	4	3	1
Percent	73.3%	13.3%	10%	3.3%

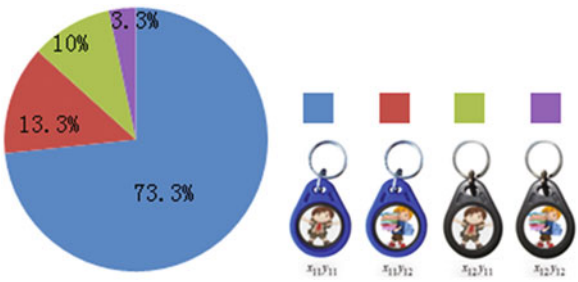


Fig. 9 Pie chart of statistical result

4.3 Evaluation and Analysis

We conduct statistical analysis of the recommendation scheme. The recommendation schemes are displayed on the monitor one by one, so that the corresponding subjects made subjective choices. Each of the subjects chose the one they were most satisfied with. Thirty of the most satisfactory scheme were chosen by 30 subjects.

We made a statistical analysis of the experimental results. The statistical results are shown in Table 3 and Fig. 9. The amount of subjects in the subjective selection scheme $x_{t1}y_{t1}$ was 22, accounting for 73.3% of the total subjects; The amount of subjects in the subjective selection scheme $x_{t1}y_{t2}$ was 4, accounting for 13.3% of the total subjects; The amount of subjects in the subjective selection scheme $x_{t2}y_{t1}$ was 3, accounting for 10% of the total subjects; The amount of subjects in the subjective selection scheme $x_{t2}y_{t2}$ was 1, accounting for 3.3% of the total subjects.

5 Conclusion

From the experimental results, we can see that the scheme $x_{t1}y_{t1}$ is the most preferred option through eye tracking technology. The probability of being picked in a repeat experiment is 73.3%, so it can be concluded that: The accuracy of using eye-tracking technology alone to quickly extract customer demand information is too low. Therefore, it is not suitable to directly use the eye tracking technology as Human-Computer Interface to extract customer demand information. Because the experiment could not select more subjects to participate in the experiment, only 30

subjects were selected, and the experimental results had certain limitations. Although eye-tracking technology is not suitable for independent use as human-computer interface, it still has certain credibility when extracting subjective demand information. Intelligent and rapid extraction of human needs information is a hot research direction. In the future research, we can combine eye-tracking technology with EEG technology to study people's real thoughts and extract their real intentions [18].

References

1. Wang K (2015) Intelligent Predictive Maintenance (IPdM) system—Industry 4.0 scenario, www.witpress.com, ISSN 1743-3533 (online) https://doi.org/10.2495/IWAMA150301
2. Wang J, Fan G, Yan F et al (2016) Research on initiative scheduling mode for a physical internet-based manufacturing system. Int J Adv Manuf Technol 84:47–58
3. Bogers M, Hadar R, Bilberg A (2016) Additive manufacturing for consumer-centric business models. Technol Forecast Soc Chang 102:225–239
4. Viviani M, Bennani N, Egyed-Zsigmond E (2016) G-Profile: a hybrid solution for extended identity management in the field of personalized service provision. Inf Resour Manage J 25 (3):61–77
5. Gong DH, Wang ZJ (2014) Evaluation of enterprise leader's qualities that affect the survival and development of enterprises. Appl Mech Mater 687–691:4622–4630
6. Zhang Y, Li Y, Wang L, Wang C et al (2016) The analysis and optimization of personalized customization model of crowdsourcing based on GSPN. Int J Multimedia Ubiquitous Eng 11 (6):411–426
7. Mathesoulek K, Slevitch L, Dallinger I (2015) Applying mixed methods to identify what drives quick service restaurant's customer satisfaction at the unit-level. Int J Hospitality Manage 50:46–54
8. Pierantonelli M, Perna A, Gregori GL (2015) Interaction between firms in new product development. Behav Brain Res 282(144):125–132
9. Kulikova NN, Smolentsev VM, Tikhonov AI et al (2016) Planning of technological development of new products and its impact on the economic performance of the enterprise. Int J Econ Financ Issues 6(S8):213–219
10. Kountouriotis G (2012) Gaze direction and visual information when steering. University of Leeds, http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.581790
11. Harris JM, Wiggins MW (2008) Evaluating cognitive competence: does eye movement behavior represent the missing piece of the puzzle? Human Factors Ergon Soc Meet 52 (26):2077–2081
12. Lemonnier S, Brémond R, Baccino T (2010) Discriminating cognitive processes with eye movements in a decision-making driving task. J Eye Mov Res 7(4):1–14
13. Boraston Z, Blakemore SJ (2007) The application of eye-tracking technology in the study of autism. J Physiol 581(3):893–898
14. Lai ML, Tsai MJ, Yang FY et al (2013) A review of using eye-tracking technology in exploring learning from 2000 to 2012. Educ Res Rev 10(4):90–115
15. Kuo FY, Hsu CW, Day RF (2010) An exploratory study of cognitive effort involved in decision under framing—an application of the eye-tracking technology. Decis Support Syst 48(1):81–91

16. Liu HC, Lai ML, Chuang HH (2011) Using eye-tracking technology to investigate the redundant effect of multimedia web pages on viewers' cognitive processes. Comput Hum Behav 27(6):2410–2417
17. Baptista PM, Mercadante MT, Macedo EC et al (2006) Cognitive performance in Rett syndrome girls: a pilot study using eyetracking technology. J Intellect Disabil Res 50(9): 662–666
18. Wang Y (2015) Introduction of neural operations management—a product design perspective, www.witpress.com, ISSN 1743-3533 (online) https://doi.org/10.2495/IWAMA150491

Big Data Analysis in Click Prediction

Lapo Chirici and Kesheng Wang

Abstract Discovering behavioral pattern within huge amount of data is one of the paramount applications in digital technologies. Beyond smart manufacturing and Industry 4.0, one of the most profitable and cutting-edge market is represented nowadays by the user clicks' forecasting on online content distribution platforms. Several algorithms able to assign priority on the content to be displayed have already been released. In the present paper a real case is discussed with the aim to cross-check data arising from different sources and to extract a valid model able to optimize the rightest combination of contents to target-users more likely to respond. In order to accomplish that, a multilayer sampling, that took into account of geographical and temporal segmentations, has been performed with the aim to preserve the data integrity and to execute more easily the operations at the same time.

Keywords Big data · Decision support · Regression model · Pattern discovery

1 Introduction

Outbrain is a US based advertising company that provides a widget to recommend set of multimedia contents—like articles, photos, videos—to users viewing specific news sites. Premium publishers use Outbrain to boost leads to their websites, promoting content on host sites (e.g. www.CNN.com), and paying Outbrain to be promoted. The promoted content appears as a on-page news recommendation ("You May Also Like…"), placed as a set of customized articles through a user' behavioural targeting algorithm. Once a click on a external link happens, the

L. Chirici (✉)
Università di Pisa, Pisa, Italy
e-mail: lapochirici@gmail.com

K. Wang
Norwegian University of Science and Technology, Trondheim, Norway
e-mail: kesheng.wang@ntnu.no

K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_66

publisher is charged through a PPC model ("Pay-Per-Click") and the revenue generated is split in two (half to the hosting website, half to Outbrain).

More than 35,000 website host regularly 250 billion circa of Outbrain recommendations, producing over 15 billion page views per month in average. The penetration in US countries is estimated as 87%. The strain here described aims to point out an alternative way to put in relationship data in terms of time slots, clicks and platforms used. The model implemented strives also to suggest a most profitable projection for the ads considering the user parameters, the ratio for Outbrain and the fitting pricing policy. To do so, the model advances a forecasting framework concerning the behavioural pattern in which people are more likely to click. As a consequence, Outbrain could enable action to increase its performance, both in terms of CPC optimization, and clusters' accuracy. In a general way, the core of the analysis is based on matching the most common digital marketing metrics together with the meaningful attributes (arisen by logistic regression), in order to support other perspectives to make decision [1].

Hence, if the advertisers could obtain a clearest vision on which ways they better perform, as a side effect, the outcome could be remarkable for the final readers too. In fact, as a side effect, these latter would receive only the combination of contents more aligned with their preferences, consequently reducing the wastes of impressions displayed by Outbrain to reach the number of clicks agreed with the publishers.

2 Data Processing

The dataset is provided by Kaggle (online Data Scientist community recently acquired by Google) and it is composed by 10 .csv files, including the navigation records from all over the world, collected between June 14th and 28th 2016.

2.1 Datasets Provided

The description is explained below:

- ***page_views.csv***—the number of times a page is displayed ("Page Impression" in online advertising).
- ***clicks_train.csv***—the training set that shows which cluster of ads is clicked.
- ***clicks_test.csv***—the same as clicks_train.csv, without the clicked_ad attribute.
- ***events.csv***—the contextual information about the adv has been displayed. It covers both the train and test set.
- ***promoted_content.csv***—details about the advertisers.
- ***documents_meta.csv***—information about the part of the site on which the document is displayed (e.g. the url path).

- ***documents_topics.csv***, ***documents_entities.csv***, and ***documents_categories.csv*** provide information about the content in a document, as well as Outbrain's confidence in each respective relationship.

2.2 Data Understanding

Each user in the dataset is represented by a unique id (*uuid*). A person can view a document (*document_id*), which is a web page with some content (e.g. a news article). On each document, a set of ads (*ad_id*) is displayed and each *ad* belongs to a campaign (*campaign_id*) run by an advertiser (*advertiser_id*). To protect its publisher partners, Outbrain is not releasing URLs of viewed or clicked stories, but rather anonymized documents and site identifiers (as well as other semantic issues).

Noteworthy: each set of *ad_id* is related to a cluster of 6 promoted content, not just one (e.g. a common display message). From now on, this will imply a slightly different interpretation of the metrics—as usually are—if compared with the classic online advertising ratios (especially in the evaluation of CTR) [2].

3 Data Cleaning

The first need was to skim something from the original dataset in order to highlight only the useful data for the project fulfillment.

After a preliminary analysis, the most informative files seemed to be *page_views.csv, events.csv* and *click_train.csv*.

The remaining dataset *documents_categories*, *documents_entities*, *documents_topics*, *promoted_content* and *sample_submission* have been discarded. In facts no correlation between the topics, categories and entities of the documents was specified from Kaggle Community.

3.1 Data Training

In addition to the *page_views.csv* file (29.71 GB), a more suitable version of the same dataset was also provided (around 45 MB). After some analysis on the sample, the decision has been to discard it too, since the collected information arisen from this latter were related to only two days of the period.

The remaining usable files (*events.csv* and *click_train.csv*) addressed to focus on a *join* operation to transform the tables in a multidimensional form. The intuition behind that was to represent the variation of clicks, in terms of **spatial** and **temporal** correlation (Table 1).

Table 1 Data overview

events.csv (23.120.126 record)	click_train.csv (87.141.731 record)
• display_id • uuid (user id) • document_id • timestamp (ms since 1970-01-01-1465876799998) • platform (desktop = 1, mobile = 2, tablet = 3) • geo_location (country>state>DMA)	• display_id • ad_id • clicked (1 if clicked, 0 otherwise)

A Designated Market Area (DMA) is a group of counties in the United States that are covered by a specific group of television stations. There are 210 DMAs in the United States

3.2 Preliminary Operation

The files have been joined through the attribute *dispay_id*, which is a unique key in *events.csv*. From the joined file, which represents the number of impressions,[1] it has been possible to infer these meaningful parameters:

1. the **exact URL** (*document_id*) of the web pages—containing the set of six sponsored stories—visited by the users (*uuid*)
2. the *set of* **promoted contents** (*ad_id*) that a user could have seen
3. the number of **clicks occurred** on these contents
4. **geolocation**, **timestamp** and **platform** (desktop, mobile, tablet) related to each user

The attributes *ad_id* and *document_id* are not ascribable to any semantic meaning, so only numerical analysis has been provided about them.

4 Attribute Manipulation

In order to save disk space, Outbrain has smudged some data publishing it in a "row" format. In facts, the timestamp attribute starts from the date 1970-01-01. To recover the real time of the visit, the community suggests to implement a rule that would add the number 1465876799998 to each timestamp given, and finally would delete the last three number (to set the error to thousandth of seconds).

Using Knime and the *Math Formula* node, the operations above have been executed and the resulting output has been parsed by FROM_UNIXTIME() function on Hive dashboard.

The next step was finalized to split in two date and time,[2] in order to analyze the media interaction more accurately in different moments within a day. The aim was

[1] An impression (in the context of online advertising) is when an ad is fetched from its source, and is countable (Source Wikipedia).

[2] **Date**: YYYY/MM/DD—**Time**: HH:MM:SS.

to figure the behavioral differences in terms of propensity-to-click, in relation to **daytime usage**, besides the platform and geographic distribution.

For this purpose the time attribute has been separated into five time slots:

- **morning** → from 6 am to 10 am
- **midday** → from 10 am to 4 pm
- **afternoon** → from 4 pm to 7 pm
- **evening** → from 7 pm to 00:00
- **overnight** → from 00:00 to 6 am

The way which the time has been partitioned with is inspired by the commonly used *day parting* in traditional broadcasting media (TV and Radio), with some adjustments accorded on the different web usage.

4.1 Attribute Manipulation: Geo_Location

Since the *geo_location* attribute covers the whole world, the analysis so performed might have presented some dispersion in terms of semantic consistency.

Thus the US logs has been filtered by taking off the DMA sub-attributes (within the Geo_locations), and keeping track merely of the Countries' division. With the new dataset, a *group-by* and *count* showed up 54 different Countries, since the US Military Zones were still included. Therefore, the nations signed with the countries codes "AA" (Armed Forces Americas), "AE" (Armed Forces Europe) and "AP" (Armed Forces Pacific) have been deleted because of their not-relevancy (less than 2000 in total). The updated dataset counted at the present circa 66 million records.

A new lightweight stratified sampling, executed with Knime *Partitioning* node, has taken into account about just the 5%, keeping anyhow fixed the same sharp weight in terms of proportion of the original one.

5 Data Exploration

The data exploration phase [3] was carried out with *Hue Notebook* and *Hive* for the data queries. The sample is composed by 3,343,196 records and, as a first glance, the clearest parameter is that the **80.5% of the users didn't clicked** after having viewed a sponsored content (obviously the remaining 20% did). Below, there will be shown all the queries used to compute the analysis and the relative percentages (Fig. 1).

```
select count(*) as total_record_sample from sample;
select clicked, count(*) as Total_count_click from sample group by clicked;
```

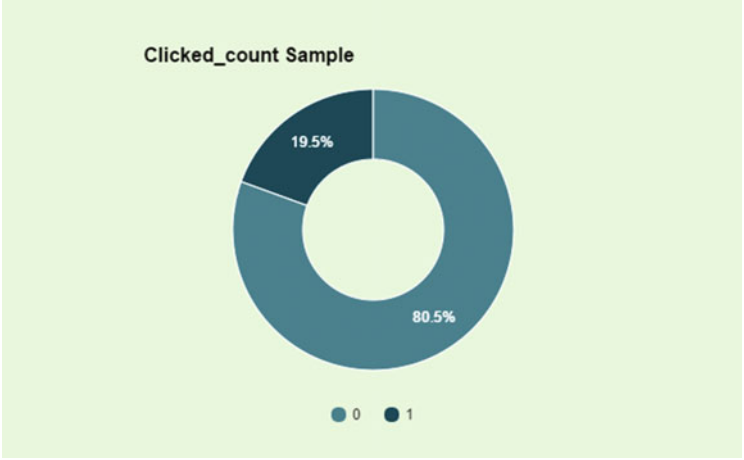


Fig. 1 Click count

6 Data Distribution

One of the most meaningful metric in order to measure the efficiency of an advertising campaign is **CTR**, **click-through rate**.[3] Since *ad_id* refers to a set of ads (six in each box), the CTR's estimation for a single content item is not that straightforward.[4] For this reason, the CTR calculated in the next part of this report refers to the whole box containing the group of six ads.

The CTR of a six-box ads is 0.19, which means 19% on the total. Potentially the average CTR of a single content could amount to 3.15% circa, but this kind of inference cannot be formally validated in absolute terms. In facts, dividing by six the CTR of the entire box, would mean assuming that the combination of advertisers inside the box would remain always fixed. And—plus—that the number of impressions for each ad should be kept constant too. However, a likely assumption of a 3% CTR represents a compelling scenario, since it is perfectly in line with the low percentage of clicked ads expressed above, and with the average metrics released by Outbrain itself (Fig. 2).[5]

```
select fasce_orarie as Fascia_Oraria, clicked as Click, count (*) as Tot_Cnt
from sample
group by fasce_orarie, clicked;
```

[3]Click-through rate (CTR) is the ratio of users who click on a specific link to the number of total users who view a page, email, or advertisement. It is commonly used to measure the success of an online advertising campaign for a particular website. **CTR = (Clicks/ Impressions) * 100**.

[4]General information about the dataset has provided here—Close-up on the content's formats is displayed here.

[5]"Generally speaking, a healthy CTR is anywhere between 0.2 and 0.25%, but is relative to varying strategies, target audiences, and content …" (Source: Outbrain).

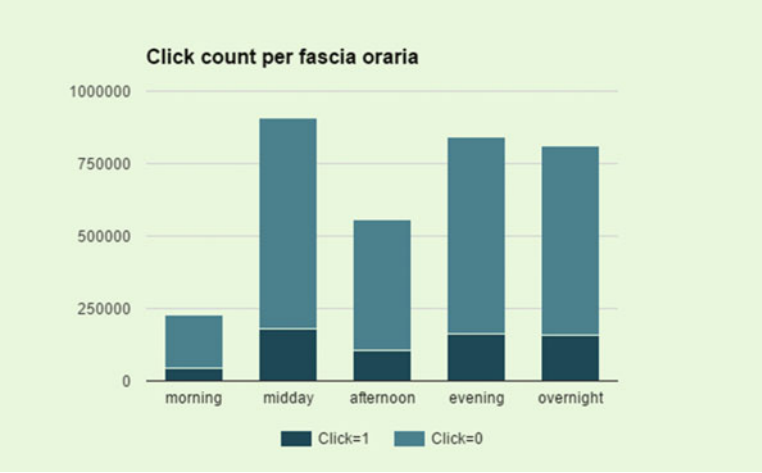


Fig. 2 Click count per time slot

6.1 Click Tendency Time-Based

As premised above, figuring out the better time to display the ads would mean for Outbrain lowering the number of impression and increasing the amount of clicks at the same time. Certainly it is important to keep in mind that the company has a **PPC business model**, that would apparently mean a not-relevancy regarding the impressions' consumption (the publisher is charged only when a click takes place) [4]. Anyway, reducing the wastes in terms of impressions could have a double significant consequence: first, it would allow Outbrain to assign the most fitting spot in terms of time slot, semantic relevancy and geographical position in relations with clicks tendency; second, since the box within the hosting platforms - as the platforms themselves (es CNN.com)—are a fixed number, placing the most clickable combination on them would increase not only the PPC performance, but also would contribute to raise the qualitative perception ("*Quality Score*") of the promoted content by targeting appropriate segments [5]. Figure 3 shows the distribution of the impressions by the *fascia_oraria* attribute.[6] Despite almost belongs to *midday*, *evening* and *overnight*, Fig. 5 reveals that the CTR is equally distributed all over the timeslot. Then the *morning* period is the more profitable for the advertisers, compared to the impression of the same timeslot and to the other period time (Fig. 4).

```
select fasce_orarie as FasciaOraria, (count(clicked = 1)/COUNT(*))*100 as
Ctr_FasciaOraria
from sample
group by fasce_orarie;
```

[6]The results number of impressions comes by the sum of clicked and not clicked set of ads.

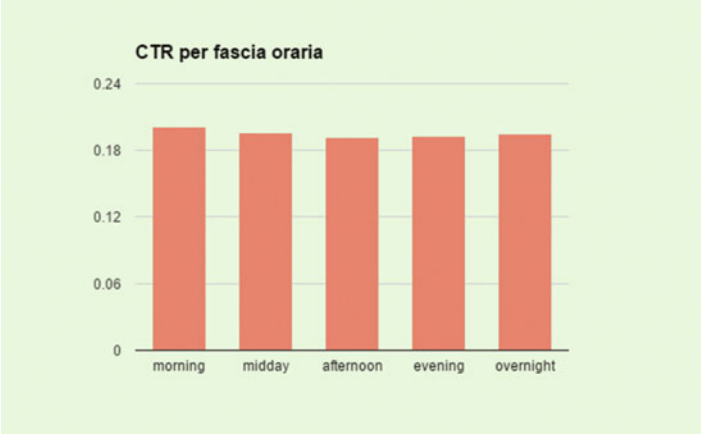


Fig. 3 CTR per time slot

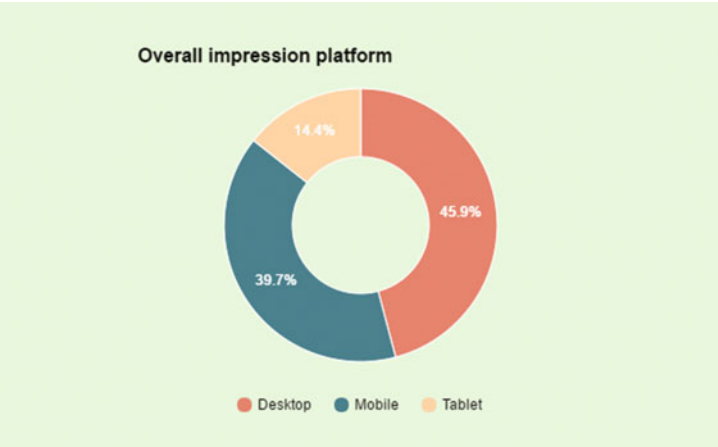


Fig. 4 Impression distribution per platform

If the company would spend more efforts in terms of content broadcasting within the morning slot, the CTR could grow more fast and—as consequence—even the Outbrain's income. Furthermore, a meaningful analysis to delve would need to take in account the areas of interest within the publishers are segmented by. So, matching the categories of promoted contents with the most inherent set of publishers, would allow Outbrain to orient investments only to "*look-alike*" targets and behaviors [6].

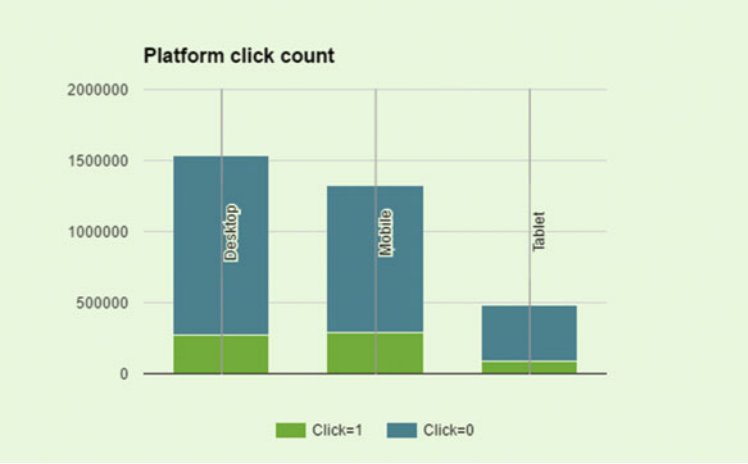


Fig. 5 Click count per platform

6.2 Data Distribution Platform-Based

Moving forward, the next operation has put in evidence which platform is more used in the different time slot.

select platform as Platform, count (*) as Impression from sample group by platform;

Mobile devices and **desktops** cover together the 85.6% of the total impressions. Besides they are also the most performing in terms of clicks. The following figure shows the total count of clicking users, grouped by platform.

select platform as Platform, count (*) as Cnt_clicked from sample where clicked = 1 group by platform;

According to the research about the most promising conditions on which an ad is more likely to be clicked, it is interesting to investigate about the distribution of platforms within the different time slots.

Here below the number of ads (clicked or not) are shown, according on the used device and grouped by time slot (Fig. 6).

Almost in all the slots, clicks come from mobile devices, especially in the morning (excepting for afternoon and evening). This supports the theory that increasing the rotation of the most performing contents in terms of CTR on the mobile platform within this time slot, Outbrain could optimize the conversions, and then the income.

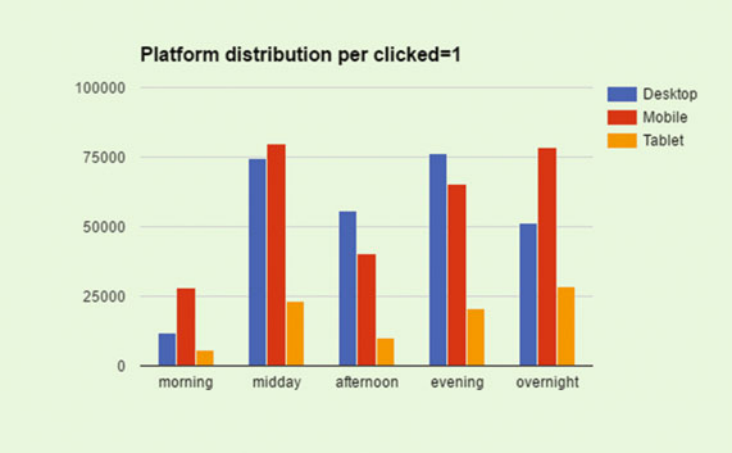


Fig. 6 Clicked ads distribution per platform

7 Data Geolocation

Regarding the geolocation, the most consistent amount of impressions are spent in **California**, **Texas**, **New York District**, **Florida** and generally in the eastern countries.

select geoloc as Geolocation, count (*) as Geolocation from sample group by geoloc;

This ratio reflects almost perfectly the correlation with other two dimensions of each state: **country's dimension** and **demographical distribution**. At the present, U.S.A. counts 325.9 million people.[7]

The following step shows the CTR distribution calculated for each country. To do so, the information provided by the *geo_location* attribute have been taken in account. Appearances aside, the most representative features are addressed by four not-that-relevant countries: **Mississippi**, **Nebraska**, **Louisiana** and **Alaska**. In facts, the highest number of reactive—and hence profitable—users have been tracked there (Fig. 7).

select geoloc as Geolocation, (count(clicked = 1)/impression)*100 as Ctr_Geolocation from sample group by geoloc;

[7]Source: Wikipedia, July 1st 2016.

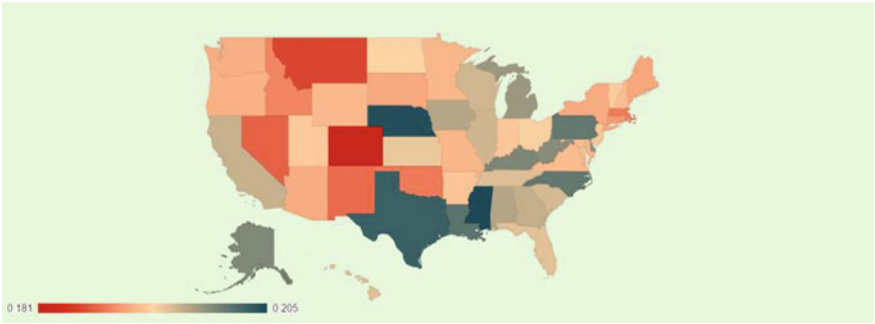


Fig. 7 CTR distribution per country

7.1 GRP's Calculation on US Population

Beside the CTR, another important metric in advertising is **GRP** (Gross Rating Point). It represents the size of an advertising campaign by a specific medium. It does not measure the size of the audience reached [7], but rather a heuristic dimension of a campaign's impact.[8] The impressions are here quantified as a percentage of the target population.[9]

$$GRPs(\%) = \frac{100 * Impressions}{Definedpopulation}$$

8 CTR Comparison with GRPs

In order to discover a correlation about the two metrics, a comparison between the CTR and GRPs has been performed too. The logic based on this ratio was to figure out if a specific relationship between **click tendency** and **impact** would exist in accordance with some advertised contents. The resulting trend showed how the two variables would seem apparently unrelated each other's. Or rather, decreasing the impact of the campaign (GRPs), the clicks tendency slightly increases in a fluctuating way. Intuitively, it is easy to infer that **quality of content matters** (Fig. 8).

[8]It is represented as a percent of the target market reached, multiplied by the exposure frequency.
[9]The number of users potentially reached by ad_id in the dataset has been considered as *defined population*.

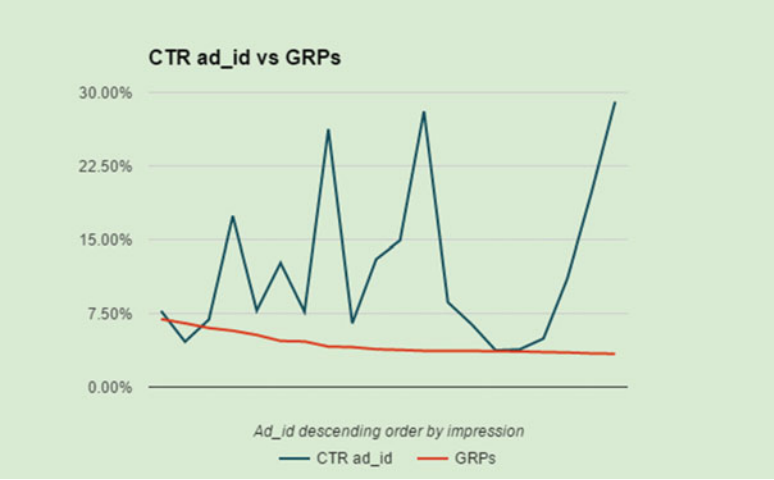


Fig. 8 CTR and GRPs comparison

9 Model Construction

In order to understand the **correlation coefficient**, a *Logistic Regression model* has been implemented with the aim to give a weight of those variables more likely to be connected with the click occurrences.

Logistic regression or *logit model* is a type of regression, considered also as a classification model, where the dependent variable is categorical [8].

9.1 Logistic Regression

In the specific case of Outbrain, the attribute involved is *clicked,* which is a binary value y(i) ∈ {0,1}, and in logit regression the estimation of the dependent variable is interpreted as a probability.

The linear probability model (Formule 1) produces an inefficient output, thus it has been necessary to transform the model in order to have the estimated probabilities in the range [0, 1]. The logistic function is graphical represented by a *sigmoid curve.*

$$\Pr(Y_i = 1|X_i) = \alpha + \beta X_i + \varepsilon_i \qquad \Pr(Y_i = 1|X_i) = \frac{\exp(\beta X_i)}{1 + \exp(\beta X_i)} = \frac{1}{1 + e^{-\beta X}} \quad (1)$$

Formule 1 − Linear probability Formule 2 − Estimated probabilities

The logit model uses the ratio between the complementary probability of the logistics function called *odd*, and the *odd logarithm* is denoted as *logit*:

$$\frac{\Pr(Y=1|X)}{\Pr(Y=0|X)} = \frac{P(y|x)}{1-P(y|x)}$$
$$= \frac{\exp(\beta X_i)}{1+\exp(\beta X_i)} \Big/ 1 - \frac{\exp(\beta X_i)}{1+\exp(\beta X_i)} \quad Logit(P_i) = \log\left(\frac{P(x_i)}{1-P(x_i)}\right) = \beta X_i + \varepsilon_i \qquad (3)$$
$$= \exp(\beta X_i).$$

Formule 3−Odd Formule 4−Logit

The magnitude of the regression coefficient *b* (estimation) denotes the strength of relationship between the variables due. This value range is between $-\infty$ and $+\infty$ and is centered on the zero. Once estimated a beta coefficient, it is useful to understand the model starting from the estimated value of the *odds* and *odds ratios.* Logit coefficients are in log-odds units and cannot be read as regular OLS coefficients. To interpret them it must estimate the predicted probabilities of Y = 1.

10 Model Validation

In order to run the regression model, a discretization of the nominal attributes (*geolocation* and *fascia_oraria*) was needed. Each distinct value has been translated in a numerical *id*, chosen starting from 1. The attribute to be predicted has been placed in the first position. It has been also provided a test set obtained as a stratified sample from 1% of the original csv.

10.1 Logit Model Construction

The first logistic regression model has been run on these attributes: *CTR*, *GRPs*, *platform*, *geolocation* and *time*. Since the coefficients related both to geolocation and platform were resulting extremely low (around 0.01), the most promising idea has been to discard them in the next operations. Moreover, the GRPs attribute already included the geographic information, so no patterns were getting lost.

For a similar reason—and to avoid a probable overfitting, the *ad_id* has been removed, since GRPs and CTR were highly interrelated each other and together comprised enough parameters for the final evaluation. Hence a new regression has been loaded again (Fig. 9).

The final model presents as independent variables *CTR*, *GRPs*, *Platform* and *Timeslot,* and *clicked* as dependent to be predicted in the test phase.

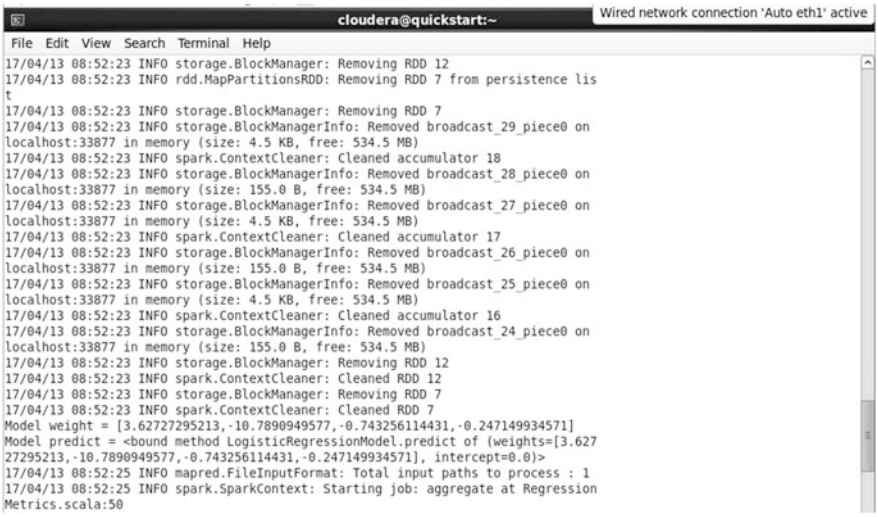


Fig. 9 Model weight and prediction

- *CTR coefficient 3.62*
- *GRPs coefficient −10.78*
- *platform −0.74*
- *timeslot −0.24*

These results reflect perfectly the inverse correlation between CTR and GRPs coefficients, following the trend described above (Fig. 10).

The accuracy has a contextual importance, since the model is able to predict a good approximation on a test set which is not directly involved in the training phase during the pattern construction.

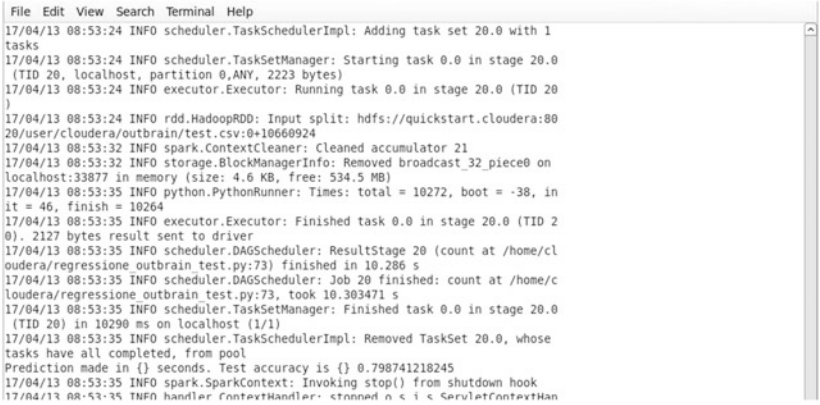


Fig. 10 Accuracy test

11 Result Exploitation

The path followed during the whole analysis led to gain a good level of accuracy, and this **can be inferred by the probability of almost 80% on clicks prediction**. This result can be considered as remarkable, especially due to its intrinsic independency by a such important attribute as *ad_id* (removed from the train set).

Thus, observing the weights distribution, a suitable path for clicked and unclicked boxes may be arisen by time slot, platform, CTR and GRPs. In particular, GRPs gives also a geographical consistency to the entire model.

12 Conclusions

The present outcome shows a high level of accuracy and it might deserve some in-depth analysis [9]. A further extension could take into account about the threshold values on which the CTR and GRPs can be manipulated, since they are turned from continuous to categorical. This would allow the model to predict with more accuracy not only the propensity for a set of contents to be clicked, but also a viable *go-to-market strategy* that considers in advance the **earnings forecasts**,[10] time and geographically based.

References

1. Kiruthika M, Rashmi J, Dixit D, Nehete A, Khodkar T (2010) Pattern discovery using association rules. Int J Adv Comput Sci Appl (IJACSA) 2(12):2011
2. Jagadeesh Chandra Bose RP, van der Aalst WMP (2009) Trace clustering based on conserved patterns: towards achieving better process models. In: Rinderle-Ma S, Sadiq S, Leymann F (eds) Business process management workshops, LNBIP, vol 43. Springer, Heidelberg, pp 170–181
3. Günther CW, van Der Aalst WMP (2007) Fuzzy mining-adaptive process simplification based on multi-perspective metrics. In: Alonso G, Dadam P, Rosemann M (eds) Business process management, LNCS, vol 4714. Springer, Heidelberg, pp 328–343
4. Henri Briand M, Fabrice Guillet M, Patrick Gallinari M, Osmar Zaaiane M (2008) Web usage mining: contributions to intersites logs preprocessing and sequential pattern extraction with low support, World Academy of Science, Engineering and Technology 48
5. Thakare SB, Gawali SZ (2010) A effective and complete preprocessing for web usage mining. Expert Syst Appl 36(3):6635–6644
6. Daş R, Türkoğlu I (2006) Extraction of interesting patterns through association rule mining for improvement of website usability, IEEE/WIC/ACM International

[10]According to a Financial Times interview with Outbrain's managing director in Europe, the network shares "about half" of any revenue generated with partner sites. The cost per click paid is said to range from $0.15 to $0.30 while click rates are in the neighbourhood of 0.50–0.75%.

7. Mobasher B, Cooley R, Srivastava J (2000) Automatic personalization based on web usage mining. Commun ACM 43:142–151
8. Anderson R, Domingos P, Weld DS (2001) Adaptive web navigation for wireless device. In: Proceedings of the seventeenth international joint conference on artificial intelligence, pp 879–884
9. Lakshmipathy R, Mohanraj V, Senthilkumar J, Suresh Y (2009) Capturing intuition of online users using a web usage mining. In: Proceedings of 2009 IEEE international advance computing conference (IACC 2009) Patiala, India