

Zengqi Sun  
Zhidong Deng  
*Editors*

# Proceedings of 2013 Chinese Intelligent Automation Conference

Intelligent Information Processing

# Lecture Notes in Electrical Engineering

Volume 256

For further volumes:  
<http://www.springer.com/series/7818>

Zengqi Sun · Zhidong Deng  
Editors

# Proceedings of 2013 Chinese Intelligent Automation Conference

Intelligent Information Processing

 Springer

*Editors*  
Zengqi Sun  
Zhidong Deng  
Department of Computer Science  
Tsinghua University  
Beijing  
People's Republic of China

ISSN 1876-1100                      ISSN 1876-1119 (electronic)  
ISBN 978-3-642-38465-3            ISBN 978-3-642-38466-0 (eBook)  
DOI 10.1007/978-3-642-38466-0  
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013939563

© Springer-Verlag Berlin Heidelberg 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

The 2013 Chinese Intelligent Automation Conference (CIAC2013) was Sponsored by Intelligent Automation Committee, Chinese Association of Automation and organized by Yangzhou University in Yangzhou, Jiangsu Province, China; 23–25, August, 2013. The objective of CIAC2013 was to provide a platform for researchers, engineers, academicians as well as industrial professionals from all over the world to present their research results and development activities in Intelligent Control, Intelligent Information Processing and Intelligent Technology and Systems. This conference provided opportunities for the delegates to exchange new ideas and application experiences face-to-face, to establish research or business relations and to find partners for future collaboration.

We have received more than 800 papers. The topics include adaptive control, fuzzy control, neural network-based control, knowledge-based control, hybrid intelligent control, learning control, evolutionary mechanism-based control, multi-sensor integration, failure diagnosis, and reconfigurable control, etc. Engineers and researchers from academia, industry, and government can gain an inside view of new solutions combining ideas from multiple disciplines in the field of Intelligent Automation. All submitted papers have been subject to a strict peer-review process; 285 of them were selected for presentation at the conference and included in the CIAC2013 proceedings. We believe the proceedings can provide the readers a broad overview of the latest advances in the fields of Intelligent Automation.

CIAC2013 has been supported by many professors (see the name list of the committees); we would like to take this opportunity to express our sincere gratitude and highest respect to them. At the same time, we also express our sincere thanks for the support of every delegate.

Zengqi Sun  
Chair of CIAC2013

# Committees

## Honorary Chairs

Yanda Li, Tsinghua University, China

Bo Zhang, Tsinghua University, China

## General Chair

Zengqi Sun, Tsinghua University, China

## General Co-Chairs

Hongxin Wu, Beijing Institute of Control Engineering

Qidi Wu, Tongji University

Jue Wang, Institute of Automation, Chinese Academy of Science

Wenli Xu, Tsinghua University

Xingyu Wang, East China University of Science and Technology

Yixin Yin, University of Science and Technology, Beijing

Zushu Li, Chongqing University

Limin Jia, Beijing Jiaotong University

Zhidong Deng, Tsinghua University

Junping Du, Beijing University of Posts and Telecommunications

## Technical Program Committee Co-Chairs

Zhidong Deng, Tsinghua University

Yixin Yin, University of Science and Technology, Beijing

Limin Jia, Beijing Jiaotong University

Junping Du, Beijing University of Posts and Telecommunications

Tianping Zhang, Yangzhou University

**Organizing Committee Co-Chairs**

Bin Li, Yangzhou University  
Tianping Zhang, Yangzhou University  
Yun Li, Yangzhou University  
Yuequan Yang, Yangzhou University  
Hua Ye, Southeast University

# Contents

<b>1</b>	<b>3D Point Cloud Based Hybrid Maps Reconstruction for Indoor Environments</b> . . . . .	<b>1</b>
	Biao Zhang and Qixin Cao	
<b>2</b>	<b>Efficient Discriminative K-SVD for Facial Expression Recognition</b> . . . . .	<b>11</b>
	Weifeng Liu, Caifeng Song and Yanjiang Wang	
<b>3</b>	<b>An Improved RANSAC Algorithm of Color Image Stitching</b> . . .	<b>21</b>
	Weijie Huang and Xiaowei Han	
<b>4</b>	<b>Target Tracking Algorithm Based on Visual Perception Mechanism</b> . . . . .	<b>29</b>
	Peng Lu, Shilei Huang, Chi Liu, Daoren Yuan and Yafei Lou	
<b>5</b>	<b>Positioning Tropical Cyclone Center in a Single Satellite Image Using Vector Field Analysis</b> . . . . .	<b>37</b>
	Jinfeng Yang and Haoren Wang	
<b>6</b>	<b>Target Extraction Study on the Vision System of Apple Picking Robot</b> . . . . .	<b>45</b>
	XinWen Cheng and XueQiang Shi	
<b>7</b>	<b>The Research on the Method of Traffic Area Dynamic Division and Optimization</b> . . . . .	<b>53</b>
	Yirong Guo, Baotian Dong and Lei Wu	
<b>8</b>	<b>Multiclass Vehicle Detection Based on Learning Method</b> . . . . .	<b>63</b>
	Zhiming Qian, Jiakuan Yang and Lianxin Duan	



<b>9</b>	<b>An Object-Level Approach Improved by Quadtree to Dynamic Monitoring of Mining Area Expansion . . . . .</b>	<b>71</b>
	Liang Huang, Yuanmin Fang, Xiaoqing Zuo and Xueqin Yu	
<b>10</b>	<b>Heuristic Optimization Algorithms for Solving MRMPT . . . . .</b>	<b>81</b>
	Chunhua Meng, Hongguo Wang, Zengzhen Shao and Yanhui Ding	
<b>11</b>	<b>A Delay-Based Analysis of Multiple Bottleneck Links of End-to-End Paths in the Internet . . . . .</b>	<b>93</b>
	Jingang Liu, Wei Peng, Yonglei Yang and Zhijian Huang	
<b>12</b>	<b>Hand Segmentation Based on Skin Tone and Motion Detection with Complex Backgrounds . . . . .</b>	<b>105</b>
	Xintao Li, Can Tang, Chun Gong, Sheng Cheng and Jianwei Zhang	
<b>13</b>	<b>Local Reconstruction and Dissimilarity Preserving Semi-Supervised Dimensionality Reduction . . . . .</b>	<b>113</b>
	Feng Li, Zhengqun Wang, Zhongxia Zhou and Wei Xue	
<b>14</b>	<b>Fusion of Gray and Grads Invariant Moments for Feather Quill Crease Recognition . . . . .</b>	<b>121</b>
	Hongwei Yue, Renhuang Wang, Jinghua Zhang and Zuihong He	
<b>15</b>	<b>A New Vision Inspired Clustering Approach . . . . .</b>	<b>129</b>
	Dequan Jin and Zhili Huang	
<b>16</b>	<b>Multiclass Vehicle Tracking Based on Local Feature . . . . .</b>	<b>137</b>
	Zhiming Qian, Jiakuan Yang and Lianxin Duan	
<b>17</b>	<b>Simulating Virtual Plants Based on Genetic Algorithm and L-Systems . . . . .</b>	<b>145</b>
	Weilong Ding, Chen Hu and Yuanwei Zhu	
<b>18</b>	<b>Shrinkage Common Spatial Pattern for Feature Extraction in Brain-Computer Interface. . . . .</b>	<b>155</b>
	Yu Zhang, Jing Jin, Bei Wang and Xingyu Wang	
<b>19</b>	<b>Detection of Overtaking Vehicles in a Highway . . . . .</b>	<b>163</b>
	Song Pan and Huaping Liu	

**20 A Generalized Gamma Distributed CFAR Algorithm for Layover and Shadow Detection in InSAR Images. . . . . 173**  
 Xianxiang Qin, Huanxin Zou, Shilin Zhou and Yun Ren

**21 Error Concealment via Best Matching Block Selection Strategy . . . . . 181**  
 Chen Yao, Lijuan Hong and Yunfei Cheng

**22 Preliminary Evaluation of Classification Complexity Measures on Imbalanced Data . . . . . 189**  
 Yan Xing, Hao Cai, Yanguang Cai, Ole Hejlesen and Egon Toft

**23 A Real-Time Tracking Algorithm Based on Gray Distribution and Distance Kernel Space. . . . . 197**  
 Weixing Li, Yating Xiao, Feng Pan and Kai Zhou

**24 Laplacian Regularized D-Optimal Design for Remote Sensing Image Classification . . . . . 205**  
 Kang Liu and Xu Qian

**25 Speaker Tracking Based on Audio-Visual Fusion with Unknown Noise . . . . . 215**  
 Jie Cao, Jun Li and Wei Li

**26 The Analysis of Epidemic Disease Propagation in Competition Environment . . . . . 227**  
 Mingsheng Hu, Suimin Jia, Qiaoling Chen, Zhijuan Jia and Liu Hong

**27 An Entity Answer Ranking Method Based on MLNs . . . . . 235**  
 Fangqiong Chen, Zhengtao Yu, Jianyi Guo, Tao Shen and Yantuan Xian

**28 A Chinese Expert Name Disambiguation Approach Based on Spectral Clustering with the Expert Page-Associated Relationships . . . . . 245**  
 Wei Tian, Tao Shen, Zhengtao Yu, Jianyi Guo and Yantuan Xian

**29 The Hierarchical Heterogeneous of Parallel Computing Model Based on Method Library. . . . . 255**  
 Jibing Duan, Xiaopeng Ji, Jinye Dou and Zhiqiang Wei

**30 Research on Weakly-Supervised Entity Relation Extraction of Specific Domain Based on Entropy Minimization . . . . . 265**  
Jun Zhao, Jianyi Guo, Zhengtao Yu, Peng Chen and Cunli Mao

**31 Using Fast Sampling-Insensitive Stereo Matching for 2-D Face Recognition . . . . . 275**  
Rui Liu, Longfei Cui, Wenke Zhang and Ming Zhu

**32 The Detection Method of Printed Registration Deviations Based on Machine Vision . . . . . 283**  
Kailong Liu, Minrui Fei, Wenju Zhou and Haikuan Wang

**33 Community Discovering Based on Central Nodes of Social Networks . . . . . 291**  
Ping Fang, Fenglong Shi, Yang Chen and Wanchun Gao

**34 An Improved Force-Directed Algorithm Based on Emergence for Visualizing Complex Network . . . . . 305**  
Hongbo Li, Wenjing Geng, Yu Wu and Xian Wang

**35 Exploring Efficient Communication in Interactive Dynamic Influence Diagrams . . . . . 317**  
He Wu, Jian Luo and Le Tian

**36 A No-Reference Remote Sensing Image Quality Assessment Method Using Visual Information Fidelity Index . . . . . 325**  
Yu Shao, Fuchun Sun and Hongbo Li

**37 3D Model Feature Extraction Method Based on the Partial Physical Descriptor . . . . . 333**  
Kuansheng Zou, Haikuan Liu, Zengqiang Chen and Jianhua Zhang

**38 Leaf Classification Methods Based on SVM and SIFT . . . . . 341**  
Yida Ye

**39 Saliency Preserved Image Fusion Using Nonsubsampled Contourlet Transform . . . . . 349**  
Liang Xu, Junping Du, Qingping Li and JangMyung Lee

**40 Adaptive Wavelet Packet Filter-Bank Based Acoustic Feature for Speech Emotion Recognition . . . . . 359**  
Yue Li, Guobao Zhang and Yongming Huang

**41 A Novel Decision-Based Algorithm for Removal of Highly Corrupted Images. . . . . 367**  
 Yiyang Wang, Zhuoer Wang and Di Zhou

**42 Research and Design of Process Data Warehouse for Business Process Assessment . . . . . 377**  
 Hui Xia, Qing Yao and Fei Gao

**43 A Novel Emergency Cross-Media Information Retrieval Model . . . . . 387**  
 Lingling Zi, Junping Du, Qian Wang and Jangmyung Lee

**44 A Framework for 3D Model Acquisition from Multi-View Images . . . . . 395**  
 Chunmei Duan

**45 Vehicle Tracking Based on Nonlinear Motion Model . . . . . 403**  
 Fan Zhang, Hong Li, Kalilou Kone and Wei Zhang

**46 Detecting Pedestrian Using Motion Information and Part Detectors . . . . . 411**  
 Lingli Xu and Zhiping Zhou

**47 Tracking Algorithm Based on Joint Features. . . . . 421**  
 Xiaofeng Shi and Zhiping Zhou

**48 Distributed Audit Secure Data Aggregation for Wireless Sensor Networks . . . . . 431**  
 Zhengdao Zhang and Zhiping Zhou

**49 Multi-Expression Based Gene Expression Programming. . . . . 439**  
 Wei Deng, Pei He and Zhi Huang

**50 The Signal Processing Method of Mixed Interference Distributed Fiber-Optic Long-Distance Pipeline Leaks Detection System . . . . . 449**  
 Zhengsong Hu, Qihua Yang, Qiang Wang and Renjie Zhang

**51 Multiple Faces Tracking via Statistical Appearance Model. . . . . 459**  
 Jie Hou, Yaobin Mao and Jinsheng Sun

**52 A Fast Dictionary Training Algorithm for Single Image Super-Resolution . . . . . 469**  
 Yiliang Lv and Jiwei Liu

<b>53</b>	<b>Inner-Knuckle-Print Verification Based on Guided Image Filtering</b> . . . . .	477
	Ming Liu and Jun Yan	
<b>54</b>	<b>Face Recognition Using Sequence Combinatorics</b> . . . . .	485
	Chunhui Wang, Ankang Hu and Fenglei Han	
<b>55</b>	<b>Distinction of Breast Tissues Based on Segmented Integral Area of Frequency-Resistance Curves</b> . . . . .	493
	Chao Wang, Yiming Wei and Ruifeng Bai	
<b>56</b>	<b>Vehicle Discrimination Using a Combined Multiple Features Based on Vehicle Face</b> . . . . .	503
	Yingnan Wang, Hong Li, Clement Kipkorir Kirui and Wei Zhang	
<b>57</b>	<b>Fast SIFT Algorithm Using Recursive Gaussian Filters</b> . . . . .	513
	Zhengyuan Ye, Shouxun Liu and Xuan Wang	
<b>58</b>	<b>A Robust Linear Camera Calibration Based on Coplanar Circles</b> . . . . .	521
	Yu Cai and Yanjin Huang	
<b>59</b>	<b>A Multi-Classification Algorithm of Semi-Supervised Support Vector Data Description Based on Pairwise Constraints</b> . . . . .	531
	Ying Zhao and Guan-jun Wang	
<b>60</b>	<b>Real-Time Vehicle Classification Based on Frequency Domain Energy Spectrum</b> . . . . .	539
	Pengfei Zhang, Haijian Li, Honghui Dong, Limin Jia and Maojing Jin	
<b>61</b>	<b>A New Combination Sampling Method for Imbalanced Data</b> . . . . .	547
	Hu Li, Peng Zou, Xiang Wang and Rongze Xia	
<b>62</b>	<b>Breast Tissue Segmentation Using KFCM Algorithm on MR images</b> . . . . .	555
	Hong Song, Feifei Sun, Xiangfei Cui, Xiangbin Zhu and Qingjie Zhao	
<b>63</b>	<b>Information-Theoretic Clustering for Gaussian Mixture Model via Divergence Factorization</b> . . . . .	565
	Jiuding Duan and Yan Wang	

**64 A Method of Acceleration Applied in Symmetric-SIFT and SIFT . . . . . 575**  
 Dong Zhao, Qi Wang, Haiyan Sun and Xiaopeng Hu

**65 A Novel Method of Image Enhancement via Multi-Scale Fuzzy Membership . . . . . 583**  
 Ce Li, Yannan Zhou and Chengsu Ouyang

**66 Adaptive Region Clustering in LDA Framework for Image Segmentation . . . . . 591**  
 Xiaoru Wang, Junping Du, Shuzhe Wu and Fu Li

**67 Methods of Recognizing True and Fake Smiles by Using AU6 and AU12 in a Holistic Way . . . . . 603**  
 Pingping Wu, Wenmin Wang and Hong Liu

**68 Constrained Silhouette Based Evolutionary K-Means. . . . . 615**  
 Zhenfeng He

**69 Traffic Light Detection and Tracking Based on Euclidean Distance Transform and Local Contour Pattern . . . . . 623**  
 Zhenyang Wang, Zhidong Deng and Zhen Huang

**70 Multi-Correlation-Based Mode Decision for Multi-View Video Coding . . . . . 633**  
 Qinghong Shen, Fongsui Wang and Sidan Du

**71 Automated Discrimination of Gait Patterns Based on sEMG Recognition . . . . . 643**  
 Fei Wang, Xiao Hao, Baoxiang Zeng, Chucheng Zhou and Song Wang

**72 Hierarchical Sparse Representation for Traffic Sign Recognition. . . . . 653**  
 Yaxiang Fan, Hao Sun, Shilin Zhou and Huanxin Zou

**73 A New Sequence-Based Approach for XML Data Query . . . . . 661**  
 Wen Li, Jin Yang, Gaofeng Sun and Sen Yue

**74 Model-Based Workpiece Positioning for Robotic Fixtureless Assembly Using Parallel Monocular Vision System . . . . . 671**  
 Weiwei Yu, Mingmin Zhai and Yasheng Chen

**75 Single-Trial Identification of Motor Imagery EEG Based on HHT and SVM . . . . . 681**  
Peng Lu, Daoren Yuan, Yafei Lou, Chi Liu and Shilei Huang

**76 Robust Visual Tracking Using Incremental Sparse Representation . . . . . 691**  
Song Pan and Huaping Liu

**77 A Study on Design of Early Warning Decision Support System of Desertification . . . . . 699**  
Zhengwei Li, Jing Du, Xianyong Meng, Chen Sun and Yongqiang Liu

**78 Fast Fusion Method of TT&C Data with Multi-Routing Transmission Model . . . . . 707**  
Bin Tian, Yue Yang, Yanhui Pan and Shengjun Luo

**79 Discovery of Static Test Configuration Model and Data Model Based on TTCN-3 Test Systems . . . . . 715**  
Yongpo Liu, Shuangmei Liu, Ji Wu and Chuangye Chang

**80 A New Method of Unknown Radar Signals Sorting . . . . . 727**  
Xiaofeng Wang, Xuzhou Zhang, Runlan Tian and Xinglong Qi

**81 The Creation of Interactive Three-Dimensional Microscope Based on VRML and JavaScript . . . . . 735**  
Di Wu

**82 Research and Implementation of Face Detection System on Android Smart Phone . . . . . 743**  
Xin Li, Yumei Zhai and Xiong Li

**83 Image Mosaic Based on SURF and Results Optimization . . . . . 751**  
Tie Jiang and Guibin Zhu

**84 Improved Blind Source Separation Based on Non-Holonomic Natural Gradient Algorithm with Variable Step Size . . . . . 761**  
Ce Ji, Baocheng Tang, Kun Yang and Mingbo Sha

**85 Multi-Level Fingerprint Classification Based on Average Frequency of Ridges for Large Scale Fingerprint Database . . . . . 769**  
Xiaoqi Peng and Yunfei Zhong

**86 Two Improved Edge Coloring Algorithms for Data Migration . . . 779**  
Gangfeng Huang, Maishun Yang and Mingming Jing

**87 Medical Image Fusion Algorithm Based on the Laplace-PCA . . . 787**  
Pengtao Zhao, Gang Liu, Cen Hu, Huang Huang and Bing He

**88 Filter Parameter Estimation in Non-local Means Algorithm . . . . . 795**  
Hong-jun Li, Wei Hu, Zheng-guang Xie and Yan Yan

**89 Image Fusion Using Compressed Sensing in Nonsubsampled  
Contourlet Transform Domain . . . . . 803**  
Fu Liu

**90 The Reverse Loop Subdivision Algorithm on Approximate  
Minimum Error . . . . . 811**  
Boning Ma, Longxing Kong, Xiaoan Tang and Gangyao Kuang

**91 A Hybrid Algorithm Based on PBIL Algorithm and Zooming  
Algorithm and Its Convergence Proof . . . . . 819**  
Gaopeng Wang

**92 Image Intensity Correction Under Illumination Changes  
Based on Mixture Gaussian Model . . . . . 831**  
Yanxiang Han, Zhisheng Zhang, Lei Zhang, Ping Chen  
and Fei Hao

**93 Maneuver Target Detection Method with Iterative Endpoint  
Fitting Assisted . . . . . 839**  
Zhangsong Shi and Zhonghong Wu

**94 Multi-Camera Tracking via Online Discriminative Feature  
and Multi-Cue MRF . . . . . 847**  
Jianyong Wang, Feng Chen, Jianwu Dong  
and Dingcheng Feng



# Chapter 1

## 3D Point Cloud Based Hybrid Maps Reconstruction for Indoor Environments

Biao Zhang and Qixin Cao

**Abstract** In this article we investigate the problem of constructing a useful 3D hybrid map for both human being and service robots in the indoor environments. The objects in our laboratory include different tables, shelves, and pillar, which are of great importance for indoor service robot. We detail the components of our map building system and explain the essential techniques. The environment is detected in 3D point clouds, after sophisticated methods operating on point cloud data removing noise points and down sampling the data, we segment the data into different clusters, estimate the posture for clusters that can be recognized from library and replace it with VRML model we built in advance, then reconstruct surface for which cannot be recognized. Finally the preliminary hybrid maps are represented with the form of point cloud, VRML model and triangular meshes in 3DMapEditor.

**Keywords** 3D Point cloud · Hybrid map · Indoor environment

### 1.1 Introduction

Autonomous service robot are playing an increasingly important role such as moving objects and cleaning up in our everyday life, as a result they must have a detailed perception of the indoor environments: the position of different objects, the length and width of the wall and other important information.

Map is essential for these service robots designed to navigate around a space with some persistent memory of the features of that space. However, high-quality maps for robots may not be very useful to human being, people want good visual

---

B. Zhang (✉) · Q. Cao  
Research Institute of Robotics, Shanghai Jiao Tong University, No 800 DongChuan Road,  
Min Hang, Shanghai City, China  
e-mail: zzb00zzb@sjtu.edu.cn

effects but robot need more location information. Our goal is to construct a hybrid map useful for both robot and people, this way it can simplify man-robot communication and save a lot of time. For example, [1] represents a navigation system based on hybrid map for intelligent wheelchair.

We approach the map reconstruction problem by designing a structured system, then we attach the system with a large database (knowledge library) prepared in advance, in the database there are both VRML models and point cloud model feature of objects in our building, VRML models and point cloud model are linked correspondingly. In actual experiments we first collect point cloud data, after some processing work we compare the segmented point clusters with objects in library, then we estimate the posture for those objects we can recognize, for those clusters cannot be recognized we reconstruct their surface, the final map can be imported into 3DMapEditor for better visualization effect without losing important information for robot navigation.

The remainder of the paper is structured as follows. The next section briefly describes an overview of related work, followed by an architecture of our system in Sect. 1.3. In Sect. 1.4, we present the preparation of point cloud data. Section 1.5 presents the results of our experiment. Finally, we conclude and plan for future work.

## 1.2 Related Works

For many years the field of map building with mobile robot platform have attracted lots of researchers, many efforts have been made. But most of them focus on localization and navigation using 2D map [2, 3], a few researchers develop algorithm for 3D point cloud processing [4, 5]. Our system is the result of combination of different algorithms and publications.

In [4], Rusu et al. investigates the problem of acquiring 3D object maps of indoor household environments, in particular kitchens. The objects modeled in these maps include cupboards, tables, drawers and shelves. Rusu et al. also investigated semantic labeling of planar surfaces in indoor environments in [5]. Their proposed approach includes a processing pipeline, including geometric mapping and learning, for processing large input datasets and for extracting relevant objects useful for a personal robotic assistant. In our approach, we improve this type of approach using larger scale scanned scenes that include multiple point clouds taken from different positions.

In [6], Alexander et al. presents an extension to their feature based mapping technique that includes information about the locations of horizontal surfaces such as tables, shelves, or counters in the map. Their preliminary results are presented in the form of a feature based map augmented with a set of 3D point clouds. We improved the hybrid map into a map containing three kinds of data structure, including point cloud, triangular mesh and VRML models.

In Dr. Martin Magnusson's doctoral thesis [7], The Normal Distribution Transform (NDT) algorithm is explained in detail, it can serve as a registration algorithm that uses standard optimization techniques applied to statistical models of 3D points to determine the most probable registration between two point clouds. We employ this algorithm to determine a rigid transformation between point cloud data sets gathered from different positions in our lab. [7] classifies 3D data from a laser sensor into walls, floor, ceiling, and doors, but their segmentation scheme are very limited, we also improve it in the segmentation section.

### 1.3 System Overview

An overview of our reconstruction system is given in this section, mainly including on-line process and off-line process, as well as outlier removal, point cloud registration and segmentation, feature extraction and other components.

A more detailed diagram of the system description can be seen in Fig. 1.1. The whole system is built on PCL (Point Cloud Library, an open project for 2D/3D image and point cloud processing), 3DMapEditor (Developed by SJTU in China, developed for 3D display and simulation) and other open source tools.

The first step of processing point cloud data and constructing 3D map is registration and segmentation, after that we can match different parts of point cloud with models in our libraries built previously, the second step is to calculate the position and rotation of all the objects, with all the information we got from steps above it would be simple to replace the whole point cloud with 3D hybrid map (Fig. 1.2).

The hybrid map refers to the combination of different data structure in the map, which indicates points, triangle meshes, 3D models, geometry shapes, and so on. Different tasks can require proper data structure from this map. For example, 3D collision detection may need triangular meshes, and object recognition requires geometry feature of point cloud. In our implementation the hybrid model map is formed by 3 different types of maps:

- (1) VRML model. We build 3D VRML model for featured furniture in our lab, including different tables, pillars, and some other robots. The scale and structure is totally the same with real objects, once a point cluster is recognized as a specific object, we can replace the point cluster with our VRML model.
- (2) Triangular meshes, used for 3D collision detection, also it shows a better visual effect than point cloud, as shown in Fig. 1.3.
- (3) Point cloud, used for those environment parts like ceiling, wall, and floor. These parts are thought of as being static or unmovable, which indicates that they are rarely manipulated or changed. However, they provide us a lots of useful information for indoor robots completing tasks.

In the following sections we will explain the key technique used when building hybrid map in detail.

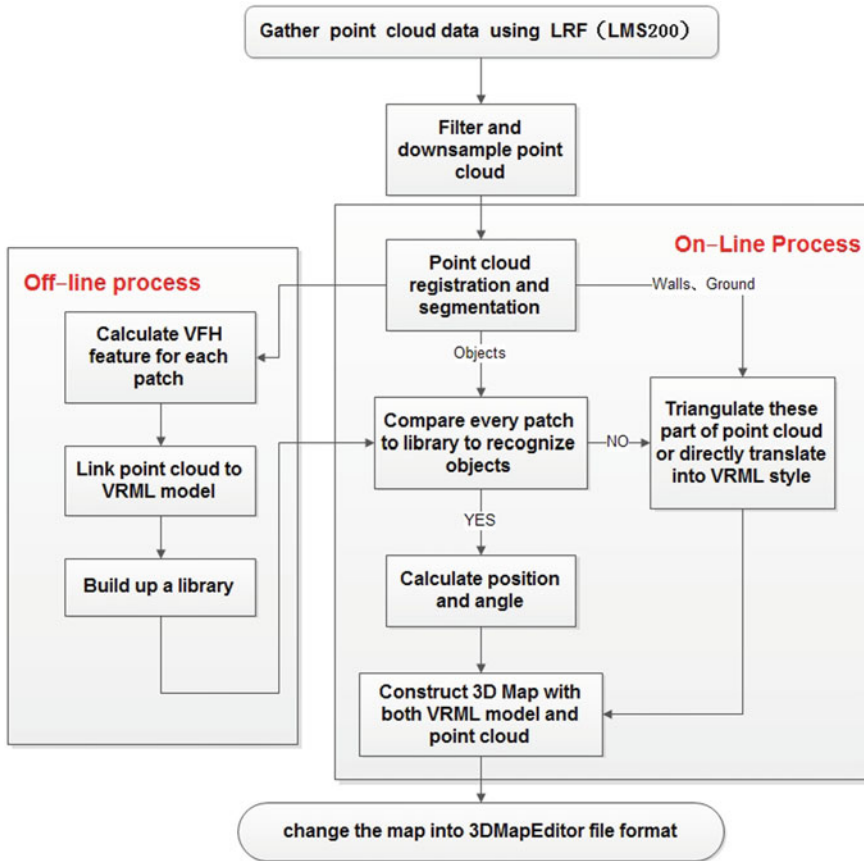


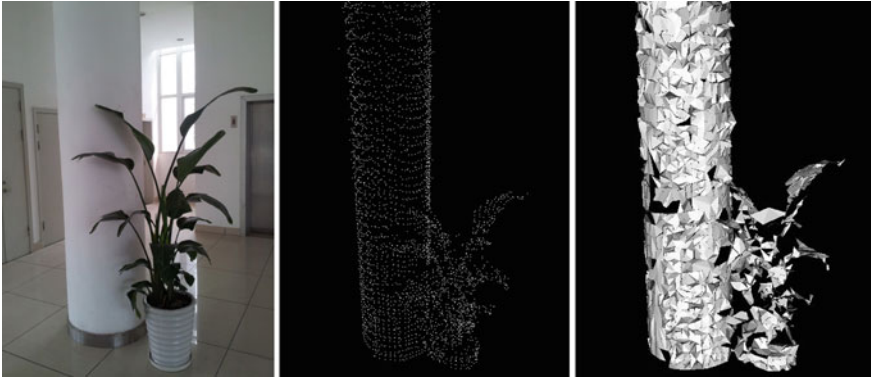
Fig. 1.1 The main frame of our map reconstruction system



Fig. 1.2 The relationship between point cloud model and VRML model, *left* point cloud; *right* VRML model

### 1.4 Point Cloud Registration and Segmentation

In this section, we describe our approach for 3D point clouds registration, as well as for segmenting 3D point cloud into independent clusters.



**Fig. 1.3** From *left to right* the real scene, point cloud cluster, and corresponding triangular meshes

### 1.4.1 Registration

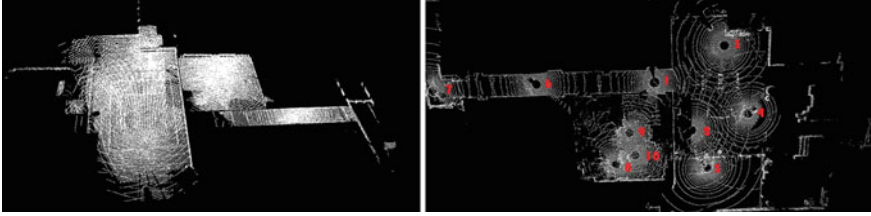
Registration is the technique of combining several datasets into a global consistent mode, its goal is to find the relative positions and orientations of the separately acquired views in a global coordinate framework. The main idea is to identify corresponding points between the data sets and find a transformation that minimizes the distance between corresponding points.

The specific description of our robot platform can be checked in [Sect. 1.5](#), when we manually drive the robot through our lab environment, we periodically stop the robot to take 3D scans using our tilting laser scanner LMS200, also odometry information is also logged. So from hardware system we can get a series of data set of point cloud gathered from different positions, as well as odometry information used as initial values in iteration of registration algorithm.

For every set of point cloud data acquired from different positions, we use NDT algorithm [8–10] to align them together into a single point cloud model, so that subsequent processing steps such as segmentation and object reconstruction can be applied.

### 1.4.2 Segmentation

Point cloud registration will yield a complete 3D point cloud map of indoor environment with all things together. But real environment may be filled with all kinds of objects, such as tables, desks, robots, and so on, segmentation is to tackle with the problem of separating different objects from each other, making it possible to recognize and replace objects in the following procedure, at the same time reducing processing time.



**Fig. 1.4** The two largest plane components extracted from point cloud data set, the ceiling plane is on the *left* and floor on the *right*, the positions where we gather data are marked with *red* number from 1 to 10

There are mainly two steps in this process, the first step is to separate ceiling, floor, and walls from the rest part of the map, to do this we use the well-known RANdom SAMple Consensus (RANSAC) method [1]. Specifically in our case the ceiling contains the most points, so we first extract the largest horizontal plane which contains most inliers, then we extract the second largest plane at approximately the position of the floor. In this process, only horizontal planes (which can be distinguished inside RANSAC algorithm) are considered. If one estimated plane is not horizontal, it would be trimmed from original point cloud, then we seek for the next largest plane, this way we can save a lot of computation time. Once we found the horizontal plane, we check its average coordinate of z-axis to decide if the result is right, the extracted ceiling and floor can be seen in Fig. 1.4. Then we estimate all the vertical planes which represent walls from the rest points, before moving to step two, all the inliers of the ceiling, floor, and walls are all trimmed.

Step two aims at separating objects in the room from each other, we use Euclidean clustering method to realize it. The main thought is to use a Kd-tree structure for finding the nearest neighbors, the detailed algorithmic steps for that would be from [11]. The clustering step serves 3 purposes: (1) to remove small individual clusters with points less than threshold value (in our work the threshold is set to 800); (2) to separate multiple objects from different position which distance larger than threshold value; (3) to reduce processing time. The clusters with sufficient points are saved into separate files for further purpose. The number and position of clusters is also sent to mapping system.

We use Viewpoint Feature Histogram (VFH) feature [12] to compare objects with models in our knowledge library, in our preliminary experiment the knowledge library contains 15 different objects including desks, tables, cupboards, shelves, robots, and so on. Those clusters can be recognized will be registered to the reference model, and get the position and angle values. Replace objects from point cloud to VRML model is the final step before we get the hybrid map, the point cloud map and triangular mesh can be loaded directly without changing their position, the position and rotation of VRML models has already been calculated. Finally we save these three kinds of map according to XML file format for 3DMapEditor to load in and check out.

## 1.5 Experimental Results

In this section, we provide an overview of our robot platform, describe our data collecting process, display the environment we mapped and our result hybrid map.

### 1.5.1 Robot Platform

The robot used in our experiment is mainly composed by a modified mobile base from our soccer robot, a SICK LMS-200 for data collection, a motor used for rotating LRF to get data of  $360^\circ$ , and a laptop computer for calculation, processing and display. The robot is also equipped with a panoramic digital camera, even though not used in this work, in our future work we may integrate these two sensors (LRF and camera) together.

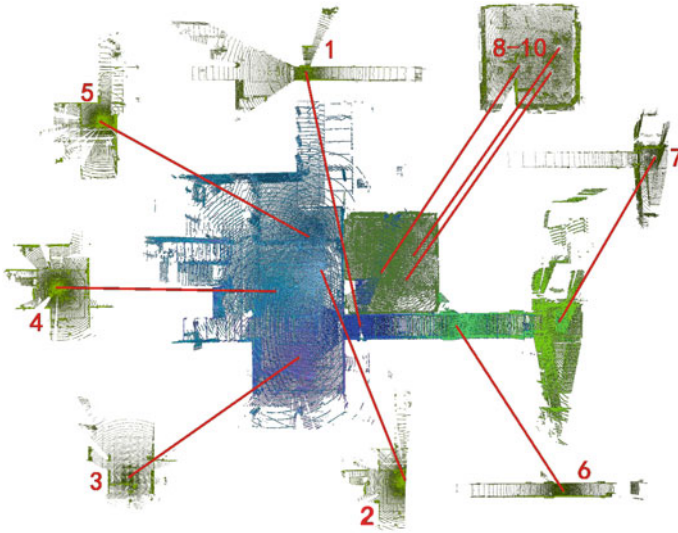
### 1.5.2 Data Collection

The robot was manually driven through our building, it was stopped at some certain places to record the odometry and point cloud data, while moving to the next place it would process the data collected from last position, saving lots of time.

The scans were taken from 10 different positions (7 of them are shown in Fig. 1.5) through this journey, position 1, 2, 6 were located in the corridor while position 2–5 were chosen in the hall, position 8–10 were chosen in a room. We tried to make scan spots distribute evenly so that different objects would end with the same detail, but still some objects have denser coverage than other areas. The elementary point cloud map after registration is displayed in Fig. 1.6.

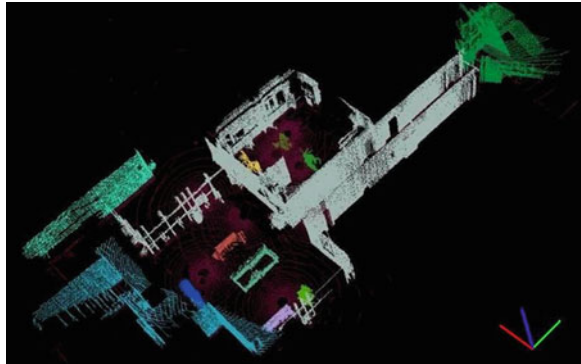


**Fig. 1.5** The layout of our experiment environment, the positions where we gather data are marked with *red spots* and are numbered from 1 to 7



**Fig. 1.6** The registration outcome of point data sets from 10 positions. The complete point cloud map is in the middle, with *red lines* lead from their registered locations to their original form in the surrounding area

**Fig. 1.7** The segmentation result of point data sets, ceiling area is removed to get a clear perspective. Different clusters are rendered distinctly. For example, walls are in argent and *blue*, floor is colored in *reddish brown*

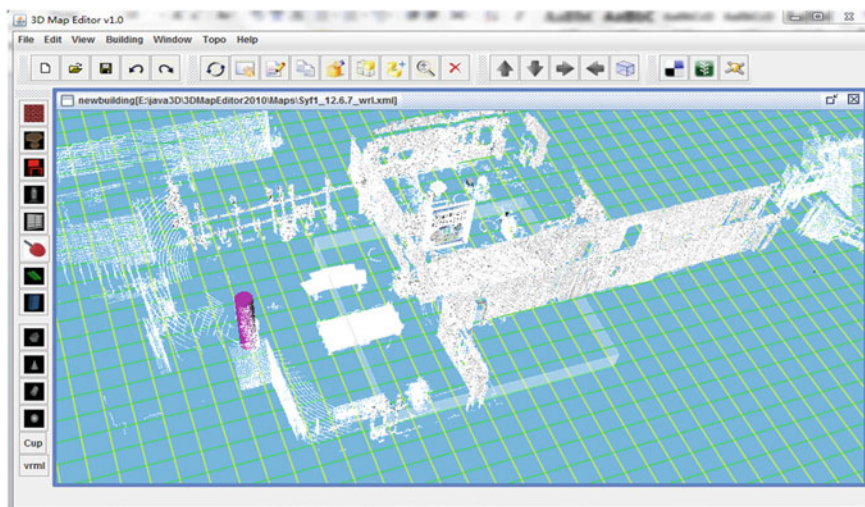


### 1.5.3 Hybrid Map

The data recorded from our robot platform contains approximately 3 million points. We tests our mapping algorithm on that point cloud data set, the intermediate result can be seen in Fig. 1.7 while the final hybrid map shown in Fig. 1.8.

In Fig. 1.7, the ceiling section is removed to achieve better perspective view, different sections are colored distinctly. The walls, floor, tables can be seen clearly from each other, almost all objects are separate successfully.





**Fig. 1.8** The final hybrid map loaded in 3DMapEditor, as is shown in the map, point cloud, triangular mesh and VRML model are all displayed together

We performed a qualitative analysis of the hybrid map. The whole point cloud map of the SiYuan building's 1st floor is more than 100 MB, after down sampling the map become a little more than 31 MB without apparent reduce in accuracy. After constructing hybrid map, the final result map is approximately 10 MB, which can be used in most mobile robots with moderate memory, and also achieves a better visual effect.

## 1.6 Conclusions

We have presented a comprehensive system for constructing 3D hybrid map based on 3D point clouds and explained the key methods used to achieve that. Our hybrid map contains 3 components: (1) a point cloud map which contains the fixed parts of indoor environment with pragmatic value (such as walls, ceilings and floor). (2) a VRML model map which is composed by the VRML model built in advance. (3) a triangular mesh map. The point cloud map is built through extracting planar regions in the original dataset, and provide pragmatic information for mobile robots. After that step, point cloud clusters will be compared with models in our knowledge base, if a point cloud cluster is recognized as an object, its pose will be estimated and it will be replaced with VRML model. Triangular mesh serves for 3D collision detection and path planning.

**Acknowledgments** This work is supported by National Natural Science Foundation of China (Grant No. 61273331), National 863 Key Program on Advanced Manufacturing Technology Field of China (Grant No. 2012AA041403) and Yaskawa Electric Corporation.

## References

1. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 24:381–395
2. Biber P, Andreasson H, Duckett T, Schilling A (2004) 3D modeling of indoor environments by a mobile robot with a laser scanner and panoramic camera. In: *IEEE/RSJ international conference on intelligent robots and systems*
3. Weingarten J, Gruener G, Siegwart R (2003) A fast and robust 3d feature extraction algorithm for structured environment reconstruction. In: *International conference on advanced robotics*
4. Rusu RB, Marton ZC, Blodow N, Dolha M, Beetz M (2008) Towards 3D point cloud based object maps for household environments. *Robotics Auton Syst J (Special Issue on Semantic Knowledge)* 56:927–941
5. Rusu RB, Marton ZC, Blodow N, Holzbach A, Beetz M (2009) Model-based and learned semantic object labeling in 3D point cloud maps of 3D point cloud based hybrid maps reconstruction for indoor environments 13 kitchen environments. In: *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS)*
6. Wang Y, Chen W, Wang J (2011) Hybrid map-based navigation for intelligent wheelchair. In: *Proceedings 2011 IEEE international conference on robotics and automation (ICRA)*, Shanghai, China, May 9–13
7. Nuechter A, Hertzberg J (2008) Towards semantic maps for mobile robots. *J Robotics Auton Syst (JRAS) (Special Issue on Semantic Knowledge in Robotics)* 56:915–926
8. Trevor AJB, Rogers JG III (2010) Tables, counters, and shelves: semantic mapping of surfaces in 3d. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS 2010) workshop on semantic mapping and autonomous knowledge acquisition (Taiwan, Oct 2010)*, IEEE/RSJ
9. Magnusson M, Lilienthal A, Duckett T (2007) Registration for autonomous mining vehicles using 3D-NDT. *J Field Robotics* 24(10):803–827
10. Steder B, Grisetti G, Burgard W (2010) Robust place recognition for 3D range data based on point features. In: *IEEE international conference on robotics and automation*, Anchorage, Alaska, USA, May 2010
11. Rusu RB (2009) Semantic 3D object maps for everyday manipulation in human living environments. *Technische Universitaet Muenchen, Germany*
12. Rusu RB, Bradski G, Thibaux R, Hsu J (2010) Fast 3D recognition and pose using the viewpoint feature histogram. In: *Proceedings of the 23rd IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Taipei, Taiwan

# Chapter 2

## Efficient Discriminative K-SVD for Facial Expression Recognition

Weifeng Liu, Caifeng Song and Yanjiang Wang

**Abstract** Dictionary learning has attracted growing attention for its prominent performance in many computer vision applications including facial expression recognition (FER). Discriminative K-SVD (D-KSVD) is one of conventional dictionary learning methods, which can effectively unify dictionary learning and classifier. However, the computation is huge when applying D-KSVD directly on Gabor features which has high dimension. To tackle this problem, we employ random projection on Gabor features and then put the reduced features into D-KSVD schema to obtain sparse representation and dictionary. To evaluate the performance, we implement the proposed method for FER on JAFFE database. We also employ support vector machine (SVM) on the sparse codes for FER. Experimental results show that the computation is reduced a lot with little performance lost.

**Keywords** Facial expression recognition · Sparse representation · K-SVD · Discriminative K-SVD · Random projection · Gabor

### 2.1 Introduction

In recent years, many new technical methods have been exploited for face recognition and facial expression recognition [1–4]. Sparse representation based classification has been performed well in facial expression recognition which

---

W. Liu (✉) · C. Song · Y. Wang  
College of Information and Control Engineering, China University of Petroleum  
(East China), Qingdao 266580, People's Republic of China  
e-mail: liuwf@upc.edu.cn

Y. Wang  
e-mail: yjwang@upc.edu.cn

exploits sparse coding to approximate an input expression image by a sparse linear combination of samples from an over-complete dictionary. In particular, dictionary learning, which aims to learn a small dictionary including few atoms from huge amounts of original information, has achieved many successful applications including audio and vision data processing [5, 6], image analysis [7–9] and certainly facial expression recognition [10]. And K-SVD [11] is state-of-the-art dictionary learning method.

Discriminative K-SVD algorithm [12, 13] extends basic K-SVD algorithm by incorporating the linear classifier into K-SVD algorithm and finally unifies the representation power and discriminate ability to train the dictionary and classifier simultaneously. D-KSVD algorithm has been proven effective and efficiency in image classification. In [10], Liu etc. utilize D-KSVD algorithm working on Gabor feature in facial expression recognition which significantly boosts the performance. However, the high dimension of facial features will cost a lot of learning time.

On the other hand, random projection (RP) [14, 15] can project original high-dimensional data onto a low-dimensional subspace using a random matrix. And Johnson–Lindenstrauss (JL) lemma [16] identifies that RP can preserve the distance between two points. In this paper, we introduce random projection as a preprocessing for feature selection and then incorporate the reduced dimensional feature into D-KSVD framework. As a result, the proposed method can effectively reduce the computation and then save the training time significantly with only a little lost of performance. Finally, we carefully construct the experiments on JAFFE dataset [25]. Experimental results demonstrate the superiority of the proposed method.

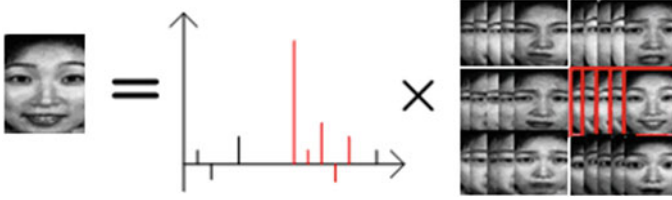
We also employ SVM on the sparse codes for FER. The verified experiments achieved the approximate results with D-KSVD algorithm.

The rest of this paper is arranged as follows. Section 2.2 introduces discriminate K-SVD algorithm in detail. Section 2.3 describes the method of random projection. Section 2.4 presents the experiment result and analysis. Finally, we conclude with discussion in Sect. 2.5.

## 2.2 Discriminative K-SVD for Facial Expression Recognition

### 2.2.1 Sparse Representation of Facial Expression Images

The basic idea of sparse representation [2–4] is using the over-complete dictionary to replace the traditional orthogonal basis and then finding the best linear combination of several atoms to represent a signal. Figure 2.1 shows the decomposition of a facial expression image using sparse representation.



**Fig. 2.1** The decomposition schematic of facial expression

As showed in Fig. 2.1, an expression can be decomposed as the linear combination of the expressions in dictionary. The coefficient corresponds to the weight of each facial expression images in dictionary.

The mathematical model can be represented as:

$$y = DX. \quad (2.1)$$

$X$  indicates the sparse coefficients. According to compressive sensing theory [17], the problem can be transformed to solve the minimum  $l_1$ -norm problem as.

$$\min \|X\|_1 \quad s.t. \ y = AX \in \mathbb{R}^m. \quad (2.2)$$

Various algorithm including L1-magic, OMP algorithm [18] etc. can be used to solve (2). Define  $\delta_i(X)$  as the coefficients of the  $i$ th class. The minimum error can be as the criteria to judge the belongings of this expression.

$$\min \xi = \|y - D\delta_i(X)\|_2 \quad (2.3)$$

### 2.2.2 Discriminative K-SVD Algorithm

Discriminative K-SVD [7, 12, 13] is the extension of K-SVD combining the representation power of K-SVD and discriminate ability of linear classifier.

K-SVD algorithm [11] tackles the drawbacks of sparse representation through learning a small-scale dictionary from the given training samples and preserving the representation power of the original dictionary. The algorithm can be achieved by the followed problem.

$$\min_{D,X} \|Y - DX\|_2^2 \quad s.t. \ \|X\|_0 \leq T \quad (2.4)$$

where  $Y$  denote the training samples and  $T$  is a fixed sparsity factor. The algorithm works well in image denosing and reconstruction. But it is restricted in image classification without considering discrimination ability.

A linear predictive classifier will be introduced to make the dictionary optimal for classification. The linear classifier  $Q(X, W, a) = W^T X + a$  can be replaced by the following optimal problem:

$$[W, a] = \arg \min_{W, a} \|Q - W^T X - a\|^2 + r\|W\|^2 \quad (2.5)$$

Set  $a = 0$ . An objective function for learning a dictionary with representation power and classification ability can be defined as the following optimal problem:

$$\min_{D, W, X} \|Y - DX\|_2 + \beta\|Q - W^T X\|_2 + r\|W\|_2 \quad s.t. \|X\|_0 \leq T \quad (2.6)$$

In order to learn the dictionary  $D$  and  $W$  simultaneously, drop the regularization penalty term  $r\|W\|_2$  and convert the problem into the following equation:

$$\min_{D, W, X} \left\| \begin{pmatrix} Y \\ \sqrt{\beta}Q \end{pmatrix} - \begin{pmatrix} D \\ \sqrt{\beta}W \end{pmatrix} X \right\|_2 \quad s.t. \|X\|_0 \leq T \quad (2.7)$$

where  $Q$  is label information. Set  $\beta = 1$ , (2.7) can be converted to:

$$\min \|\tilde{Y} - \sum \tilde{d}_i \tilde{x}_i\| = \min_{\tilde{d}_k, \tilde{x}_k} \left\| \left( \tilde{Y} - \sum_{k \neq i} \tilde{d}_i \tilde{x}_i \right) - \tilde{d}_k \tilde{x}_k \right\|_F = \min_{\tilde{d}_k, \tilde{x}_k} \|\tilde{E}_k - \tilde{d}_k \tilde{x}_k\|_F \quad (2.8)$$

where  $\tilde{Y} = \begin{pmatrix} Y \\ Q \end{pmatrix}$ ,  $\tilde{D} = \begin{pmatrix} D \\ W \end{pmatrix}$ ,  $\tilde{d}_k$  indicate the  $k$ th atom of the dictionary  $\tilde{D}$  and  $\tilde{x}_k$  is the corresponded coefficient.

At testing phrase, the label of the test image can be obtained through the product between linear classifier  $W$  and the sparse coefficient  $\alpha$ .

$$label = W * \alpha \quad (2.9)$$

The maximum of *label* can be viewed as the class of the test image.

### 2.3 Random Projection

There exist many dimensionality reduction algorithms [19–22] which project the data into a reduced subspace to reinforce the discriminate capability. RP (random projection) [14, 15] has been applied widely in dimension reduction. The principle of RP is very simple and easy to implement. The central idea is aroused by Johnson-Linderstrauss lemma [16]: Given an image matrix  $\Gamma$  which contains  $N$  points in  $d$ -dimensional vector space, the matrix can be projected to a lower-dimensional space while the distance between two points is preserved.

$$\tilde{\Gamma}_{k \times N} = R_{k \times d} \cdot \Gamma_{d \times N} \quad (2.10)$$

The transformed matrix  $R$  has many formats. In this paper, we select the sparse projection matrix proposed by Achlioptas [23].

$$R(r_{ij} = z) = \begin{cases} 1/6 & z = \sqrt{3} \\ 2/3 & z = 0 \\ 1/6 & z = -\sqrt{3} \end{cases} \quad (2.11)$$

This distribution reduces the computational time to calculate  $R \cdot \Gamma$ . Random vectors are sufficiently approximate to orthogonal [24], so we can use the sparse random matrix directly.

The proposed efficient D-KSVD based on random projection for facial expression recognition can be described in table:

The proposed efficient D-KSVD algorithm for FER	
Input:	The training facial images and test facial image
Output:	The label of the test facial images
Step 1	Extract the feature of facial expression images
Step 2	Reduce the dimension of the feature using random projection
Step 3	Learn the dictionary $D$ and classifier $W$ adopting D-KSVD algorithm
Step 4	Find the sparse coefficients $\alpha$ of the test sample $y$ exploit OMP algorithm
Step 5	The label of test sample can be obtained finally.
Step 6	The coefficient $X$ of step 3 and $\alpha$ of step 5 as the training sample and testing samples separately would be sent to SVM for classifying

## 2.4 Experimental Results and Analysis

A series of experiments of discriminative K-SVD with RP are performed on the JAFFE database [25]. JAFFE database contains 213 expression images in total including seven classes of expressions (Angry, Disgust, Fear, Happy, Sad, Surprise and Neutral) of ten Japanese women, which each expression has two to four images.

We use the cropped and normalized face expression images of 120\*96 pixels. The images are split into two groups which one contains 70 images as test sample, and the others as training sample.

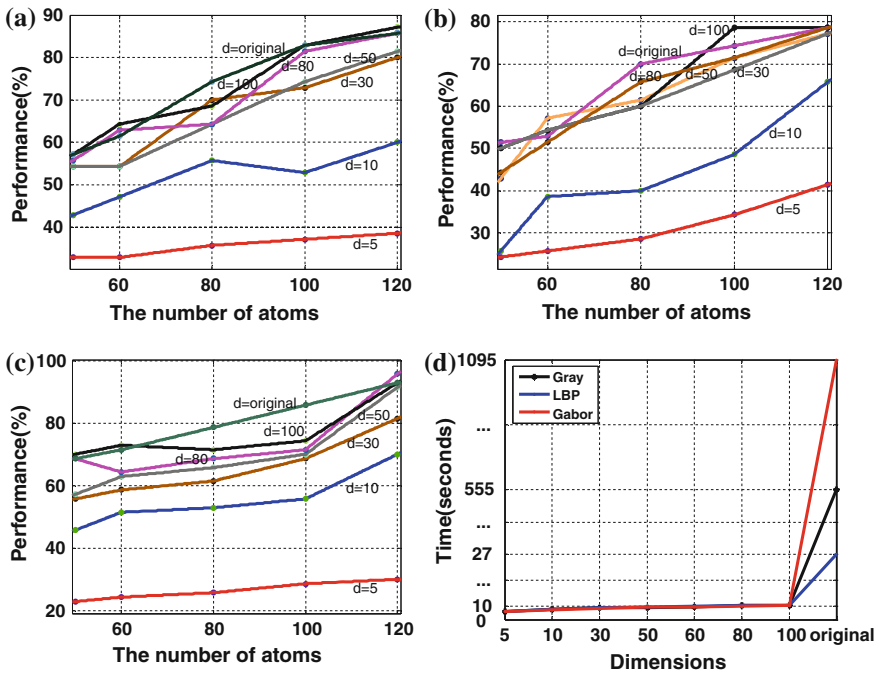
### 2.4.1 Experiments Analysis of Efficient D-KSVD Algorithm

The role of feature [26] is not neglected and various features have been applied in facial expression recognition. Three features including gray-scale feature, LBP feature [27] and Gabor feature [28] have been selected. The original dimensions are 11520, 512, 11520\*4 for gray feature, LBP and Gabor feature respectively.

Then the experiments based on discriminative K-SVD perform on the features respectively in different reduced dimensions using random projection. Each image can be separately projected into eighty-dimensions, sixty-dimensions, fifty-dimensions, thirty-dimensions, twenty-dimensions, ten-dimensions and five-dimensions. At last, the contrast of the recognition results between the original data and dimensionality reduction data is given as shown in Fig. 2.2a–c. While the comparison of training time is presented in Fig. 2.2d.

From Fig 2.2a–c, it can be seen that the recognition rates remain comparable until the dimension reduced to ten dimensional. On the other hand, in (d), the training time is decreased significantly after dimensional reduction. In addition, the performance is not only determined by the classifier, but also the extracted feature. The recognition results of D-KSVD algorithm with Gabor feature performs best in the three features.

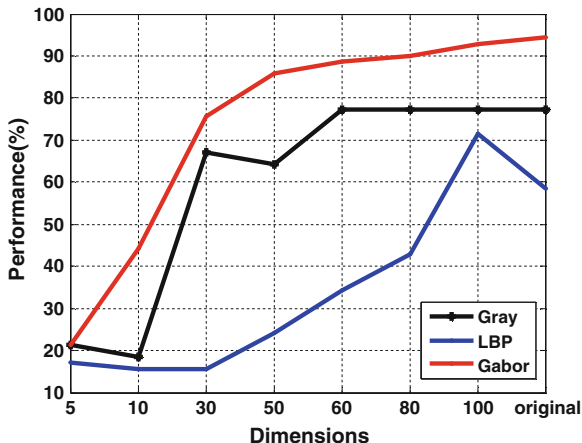
As showed in (c), On the whole, fixing the number of atoms, the results keep steady until the dimension drops to a lower number such as five dimensions. On the other hand, fixing the dimensions, the results show a rising trend with the number of atoms increasing, but it will remain unchanged basically when the atoms reaches a certain number.



**Fig. 2.2** Contrasts of performance and training time between original data and the data after dimensional reduction **a** Gray feature, **b** LBP feature, **c** Gabor feature, **d** Contrasts of training time



**Fig. 2.3** The performance of three methods using SVM in different dimensions



### 2.4.2 Verification Experiments Using SVM

The training coefficient  $X$  and the testing coefficient  $\alpha$  as the training feature and the testing feature separately would be classified using SVM algorithm. The classified accuracy would be presented in Fig. 2.3.

As shown from the table, the performance using SVM is in line with that using D-KSVD algorithm. The overall trend is declined with the reduction of dimensions. However, the accuracy after dimension reduction still remains unchanged basically if the reasonable and appropriate dimension is selected.

## 2.5 Conclusion and Future Work

An efficient discriminative K-SVD algorithm for dictionary learning is proposed for facial expression recognition. Particularly, dimensionality reduction uses random projection acted on a series of features (gray, LBP, Gabor) which are extracted from the facial expression images. Then the features after dimensionality reduction are used to train the small-size dictionary and classification simultaneously. Finally, the dictionary and classification are implemented for facial expression recognition system. Experimental results demonstrate that the training time can be greatly reduced with little performance lost after dimension reduction using RP.

**Acknowledgments** This paper is supported by the National Nature Science Foundation of China (No. 61271407), the Nature Science Foundation of Shandong Province (No. ZR2011FQ016) and the Fundamental Research Funds for the Central Universities (No. 13CX02096A).

## References

1. Song M, Tao D, Huang X, Chen C, Bu J (2012) Three-dimensional face reconstruction from a single image by a coupled RBF network. *IEEE Trans Image Process* 21(5):2887–2897
2. Wright J, Yang A, Ganesh A, Sastry S, Ma Y (2009) Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intel* 31:210–227
3. Cotter SF (2010) Sparse representation for accurate classification of corrupted and occluded facial expressions. *ICASSP* 838–841
4. Huang M, Wang Z, Ying Z (2010) A new method for facial expression recognition based on sparse representation plus LBP. *CISP* 4:1750–1754
5. Monaci G, Vandergheynst P, Sommer PT (2009) Learning bimodal structure in audio-visual data. *IEEE Trans Neural Netw* 20(12):1898–1910
6. Tian X, Tao D, Rui Y (2012) Sparse transfer learning for interactive video search reranking. *ACM TOMCCAP* 7665:269–276
7. Mairal J, Bach F, Ponce J et al. (2008) Discriminative learned dictionaries for local image analysis. *CVPR* 14(12):2091–2106
8. Protter M, Elad M (2009) Image sequence denoising via sparse and redundant representation. *IEEE Trans Image Process* 18(1):27–35
9. Zepeda J, Guillemot C, Kijak E (2011) Image compression using the iteration-tuned and aligned dictionary. *ICASSP* 1:739–796
10. Liu W, Song C, Wang Y (2012) Facial expression recognition based on discriminative dictionary learning. *ICPR* 1:1839–1842
11. Aharon M, Elad M, Bruckstein A (2006) K-SVD: an algorithm for designing over-complete dictionaries for sparse representation. *IEEE Trans Signal Process* 54(11):4311–4322
12. Pham D, Venkatesh S (2008) Joint learning and dictionary construction for pattern recognition. *CVPR* 1–8
13. Zhang Q, Li B (2010) Discriminative k-svd for dictionary learning in face recognition. *CVPR* 2691–2698
14. Majumdar A, Ward RK (2010) Robust classifier for data reduced via random projections. *IEEE Trans Syst Man, Cybern, Part B: Cybernet* 40:1359–1371
15. Sulić V, Perš J, Kristan M, Kovačič S (2010) Dimensionality reduction for distributed vision system using random projection. *ICPR* 31:380–383
16. Johnson WB, Lindenstrauss J (1984) Extensions of lipshitz mapping into hilbert space. In: *Contemporary mathematics, conference in modern analysis and probability*, vol 26, pp 189–206
17. Donoho DL (2006) Compressed sensing. *IEEE Trans Inf Theory* 52(4):1289–1306
18. Cai TT, Wang L (2011) Orthogonal matching pursuit sparse signal recovery with noise. *IEEE Trans Inf Theory* 57:4680–4688
19. Yu H, Bannamoun M (2006) 1D-PCA, 2D-PCA to n D-PCA. *ICPR* 4:181–184
20. Guan N, Tao D, Luo Z, Yuan B (2012) Online nonnegative matrix factorization with robust stochastic approximation. *IEEE Trans Neural Netw Learn Syst* 23(7):1087–1099
21. Guan N, Tao D, Luo Z, Yuan B (2012) NeNMF: An optimal gradient method for nonnegative matrix factorization. *IEEE Trans Signal Process* 60(6):2882–2898
22. Guan N, Tao D, Luo Z, Yuan B (2011) Non-negative patch alignment framework. *IEEE Trans Neural Netw* 22(8):1218–1230
23. Achlioptas D (2003) Database-friendly random projections. *J Comput Syst Sci* 66(4):671–687
24. Lin J, Gunopulos D (2003) Dimensionality reduction by random projection and latent semantic indexing. *SDM*
25. Lyons M, Akamatsu S (1998) Coding facial expressions with gabor wavelets. *AFGR* 200–205
26. Song M, Tao D, Liu Z, Li X, Zhou M (2010) Image ratio features for facial expression recognition application. *IEEE Trans Syst, Man, Cybernet, Part B* 40(3):779–788

27. Tan N, Huang L, Liu C (2008) Face recognition based on LBP and orthogonal rank-one tensor projections. ICPR
28. Tao D, Li X, Wu X, Stephen, Maybank J (2007) General tensor discriminant analysis and gabor features for gait recognition. *IEEE Trans Pattern Anal Mach Intell* 29(10):1700–1715

# Chapter 3

## An Improved RANSAC Algorithm of Color Image Stitching

Weijie Huang and Xiaowei Han

**Abstract** Image mosaic is a technique being used to stitch multiple images together to form a stitched image with higher resolution and large field of view. This paper presents an improved RANSAC algorithm of color image mosaic. Image mosaic mainly comprises two steps, namely image registration and image fusion. This algorithm comprises main modules, such as Harris corner detection, NCC rough matching, improved RANSAC exaction, estimating the projection transformation matrix, projection transformation, image smoothing. The improved RANSAC for feature points exaction based on statistical regularities can greatly support the real time of the algorithm. The experimental results show that the algorithm presented is fast and effective.

**Keywords** Image mosaic · Feature point detection · Improved RANSAC exaction · Projection transformation · Image smoothing

### 3.1 Introduction

As the images captured with the camera will contain less information than what can be observed by the human eye, image mosaic technique combines two or more images that are partly overlapped to form a high resolution image [1]. Widely used in digital video compression, motion control, virtual reality technique, remote sensing image processing [2] and medical image analysis, image mosaic technique

---

W. Huang (✉) · X. Han  
School of Information and Engineering, Shenyang University,  
Shenyang, China  
e-mail: huang1987515@126.com

X. Han  
e-mail: hxw69@163.com

has become a hotspot in photo graphics, computer vision, image processing and computer graphics.

Image mosaic mainly comprises two steps, namely image registration and image fusion. Image registration is the most difficult and critical procedure that not only extracts feature points from images, but also searches correct matching points. Generally, feature points detection includes many methods, such as Harris feature points, KLT feature points, Susan feature points, SIFT feature points, Surf points and other methods.

In recent years, researchers have done a lot in the image mosaic technology. Richard Szeliski proposes a panoramic image mosaic algorithm, based on motion, which calculates geometric transformations between images [3, 4]. The result achieves image registration and image smoothing. Document [5] puts forward a new image mosaic technology, which uses a rotation matrix to represent the relationship between the mosaic images. This method improves the clarity of the image mosaic. Compared with the foreigners, in China image mosaic technology starts late, but the development is fast. Sun Jiaguang, Liu Qiang and Qi Chi, in Tsinghua University, raise a cylindrical panorama method by tracing texture feature between adjacent frames [6, 7]. The precision of this algorithm is high but it also needs too much computation. In the image mosaic algorithms, RANSAC exaction is an important step [8], and this step is poor real-time. This paper presents an improved RANSAC algorithm based on statistical regularities [9, 10] of feature points. In this method, some mismatching points existing in no overlapping regions will be rejected [11] to support the real time.

## 3.2 Image Mosaic

### 3.2.1 Harris Feature Point Detection

The Harris operator which is based on the characteristics of the signal point extraction is proposed by C. Harris and M.J. Stephens. It has some features, such as simple calculation, uniform extraction, stable operator, and so on.

- (1) Each point on the left frame and the right frame is filtered by the improved horizontal and vertical difference operators.

$$l = \begin{bmatrix} Ix^2 & IxIy \\ IxIy & Iy^2 \end{bmatrix} \quad (3.1)$$

$$Ix^2 = Ix * Ix \quad (3.2)$$

$$Iy^2 = Iy * Iy \quad (3.3)$$

- (2) Use Gaussian filter to filter the four parameters of  $l$ , then get the new  $l$ . The discrete Gaussian function is:

$$Gauss = \exp\left(-\frac{(x^2 + y^2)}{2\sigma^2}\right) \quad (3.4)$$

(3) Calculate the tensor of feature point:

$$cim = \frac{I_x^2 * I_y^2 - (I_x I_y)^2}{I_x^2 + I_y^2} \quad (3.5)$$

(4) Finally, when the *cim* meet the threshold and the point is local maximum point, the feature point is found.

### 3.2.2 Normalized Cross Correlation Method

The feature points extracted by Harris operator may be mismatched, so it must be dealt with the matching method. Normalized cross correlation method (NCC) is a classical algorithm which calculates the cross-correlation value between the template image (Left) and the matching image (Right).

- (1) 5 neighborhoods of the feature points in two images are filtered by the smoothing operator.
- (2) By using the result of 1, a normalized cross-correlation matrix is raised. Let the size of the search image  $S$  be  $M \times M$  and the size of the template  $T$  be  $M \times N$ .  $M, N$  both are the pixels and  $M$  is larger than  $N$ . Let template  $T$  move on the image  $S$ .  $S^{i,j}$  is the area covered by template  $T$  in the image  $S$ .  $i, j$  are the coordinates of the top left corner in the  $S^{i,j}$ . The measure function measures the similarity between the search image and templates.

$$R(i,j) = \frac{\sum_{m=1}^M \sum_{n=1}^N S^{i,j}(m,n)T}{\sqrt{\sum_{m=1}^M \sum_{n=1}^N [S^{i,j}(m,n) - \bar{S}^{i,j}]^2} \times \sqrt{\sum_{m=1}^M \sum_{n=1}^N [T(m,n) - \bar{T}]^2}} \quad (3.6)$$

- (3) On basis of the normalized cross-correlation matrix, the maximum values of the feature points are calculated.
- (4) By using the result of 3, the initial matched points are found.
- (5) Cycle to calculate all the feature points.

### 3.2.3 The Improved RANSAC Exaction

Random sample consensus (RANSAC) algorithm is able to match the feature points with noise and it has good robustness. As few feature points are in no overlapping areas and there are obvious differences between no overlapping areas

and overlapping areas, the improved RANSAC exaction can easily remove the no overlapping areas based on statistical regularities of feature points.

- (1) The whole image is divided into eight pieces. Count the number of the feature points in every region. The result shows that more numbers of feature points are in the overlapping areas while fewer feature points are in no overlapping areas. The no overlapping areas are removed while the overlapping areas are preserved by the number of feature points.
- (2) In the preserved areas, four pairs of feature points are randomly selected.
- (3) Set the number of model consistent points to zero.
- (4) Select points from  $(x_1, y_1)(x_2, y_2) \dots (x_3, y_3)$  and  $(x'_1, y'_1)(x'_2, y'_2) \dots (x'_n, y'_n)$ . Point  $(x_i, y_i)$  and Point  $(x'_i, y'_i)$  make the projection transformation produce support points. According to the projection transformation model,  $a_{11}, a_{12}, b_1, a_{21}, a_{22}, b_2$  are calculated.

$$d((x_i, y_i), (x''_i, y''_i)) = \sqrt{(x_i - x''_i)^2 + (y_i - y''_i)^2} < T \quad (3.7)$$

If (3.7) is right, the number of consistent points adds one.

- (5) If the number of consistent points is greater than a given threshold value  $N_i$ (10 %), this model will be the best model. Otherwise, the process repeats from steps 1–3.

### 3.2.4 Image Smoothing

In order to eliminate the obvious edge, a linear transition method which is often used in the overlapping areas is proposed in this paper. The pixel value of the overlapping area can be calculated according to Eq. (3.8).

$$W = k \times WL + (1 - k) \times WR \quad (3.8)$$

where,  $k$  is the weighted average coefficient ( $0 < k < 1$ ). In the overlapping area of the image portion, along the left to the right  $k$  becomes 1 to 0.

$$k = \frac{d_1}{d_1 + d_2} \quad (3.9)$$

$d_1, d_2$  are the distances from the matched point to the left boundary and the right boundary in the overlapping region.

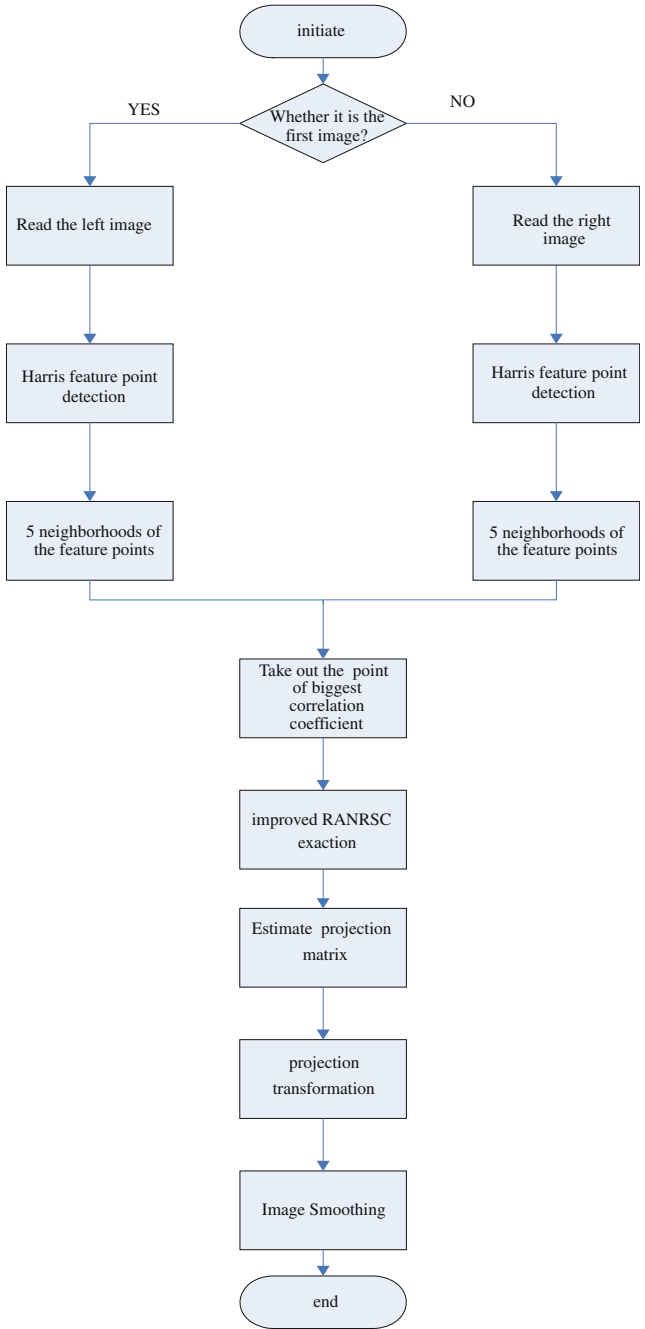


Fig. 3.1 Flowchart



### 3.3 Simulation Results Analysis

Experimental hardware platform is a PC, which contains 2.93 GHz CPU and 2.00 GB memory. The software platform is Matlab R2009a/Simulink, which is system simulation platform.

In this method, main modules are feature point detection, NCC feature points coarse matching, RANSAC feature points matching, projection transformation, image smoothing (Fig. 3.1).

Original images are shown in (Figs. 3.2, 3.3).

NCC is shown in (Fig. 3.4).

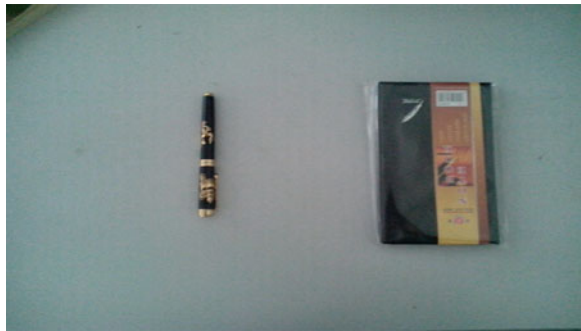
The result of improved RANSAC method is shown in (Fig. 3.5).

From the above experiment, Improved RANSAC method, which removes the outliers, can greatly support the real time. At the same time, calculating parameters of projection matrix are prepared for projection transformation. Image smoothing eliminates the verge.

The result of image smoothing is shown as follows (Fig. 3.6).

Comparing excluding outliers and including outliers are shown in (Table 3.1).

**Fig. 3.2** Left image



**Fig. 3.3** Right image



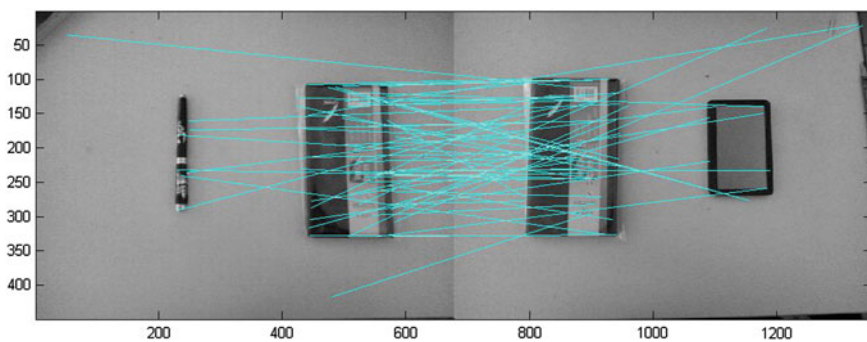


Fig. 3.4 Ncc result

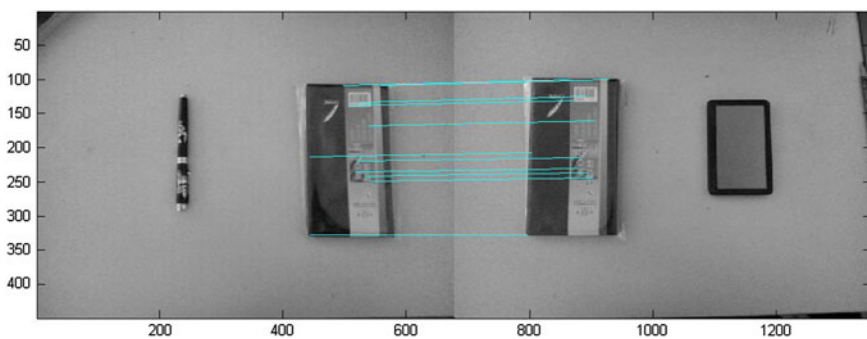
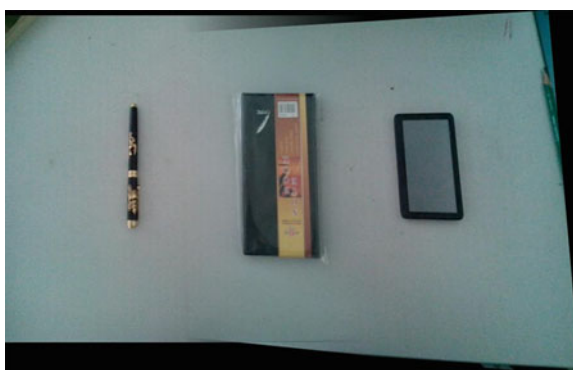


Fig. 3.5 Improved RANSAC exaction

Fig. 3.6 Image smoothing



**Table 3.1** Improvement results

	Including outliers	Excluding outliers
The number of feature points	54	42
Algorithm time	2.507894 s	0.672846 s

### 3.4 Conclusions

This paper proposes an improved RANSAC algorithm of color image mosaic. It suits different times and different camera angles. Harris corner detection, NCC angle point rough match, improved RANSAC exaction, estimating the projection transformation matrix, projection transformation, image smoothing are the main modules of the algorithm. The improved RANSAC algorithm is based on statistical regularities of feature points. In this method, some mismatching points exists in never overlapping regions will be rejected to support the real time.

### References

1. Lowe DG (2004) Distinctive image features from scale-invariant key points. *Int J Comput Vision* 60(2):91–110
2. Zhang Y (2007) *Image engineering*. Tsinghua University Press, Beijing, pp 105–107 (in Chinese)
3. Helmer EH, Ruefenacht B (2005) Cloud-free satellite image mosaics with regression trees and histogram matching. *Am Soc Photogram Remote Sens* 2005(9):1079–1089
4. Choi Y-S, Koo B-K, Lee J-H (2007) Template based image mosaics. *Lect Notes Comput Sci* 43(5):475–478
5. George P, Petros M (2008) Adaptive and constrained algorithms for inverse compositional active appearance model fitting. In: *Proceedings of the IEEE international conference on computer vision and pattern recognition (CVPR-2008)*, Anchorage, vol 1, pp 1–8
6. Jiang J, Tan R, Yang D (2009) Image splicing based on contour feature and wavelet transform image splicing. *Comput Eng* 35(3):225–229 (in Chinese)
7. Zhang Q, Liu Z, Pang Y (2008) Automatic image registration based on SUSAN algorithm. *J Surv Mapp* 32(3):245–250 (in Chinese)
8. Xi Y, Guo B, Zhan L, Yang Z (2009) Estimation affine parameters based on the image feature points. *Comput Sci* 6(36):279–281 (in Chinese)
9. Feng Y, Dai M, Sun L (2010) The optimization of automatic image mosaic. *Opt Precis Eng* 18(2):470–476 (in Chinese)
10. Guo S, Fang X, Luo B (2007) A video sequence stitching algorithm. *Comput Appl* 11(27):2786–2788 (in Chinese)
11. Kang Y, Pang C (2008) The real-time video image mosaic. *Video Signal Inf Proc* 5(38):26–28 (in Chinese)

# Chapter 4

## Target Tracking Algorithm Based on Visual Perception Mechanism

Peng Lu, Shilei Huang, Chi Liu, Daoren Yuan and Yafei Lou

**Abstract** A method based on visual perception mechanism is proposed for solving the problem of target tracking. The tracking of target can be achieved in stability. In this paper, the algorithm use neural responses as the visual features. Firstly, the receptive field of cells in primary visual cortex is obtained from natural images. Then the neurons response of background image and video image sequences can be received and calculated the difference, and the difference is compared with dynamic threshold, the target can be detected in this way. Finally, the target tracking can be realized by iterative. Many categories experiment results show that this method improve accuracy and robustness of the tracking algorithm in condition of time-real.

**Keywords** Target tracking · Visual perception · Overcomplete set · Neural responses

### 4.1 Introduction

There are a lot of target tracking methods, which are divided into region-based, feature-based, deformable template-based and model-based generally [1], Among them, the typical algorithm include Camshift [2, 3] and SIFT [4, 5], and so on.

The sparse coding model of complete basis requires the orthogonal basis functions [6]. It does not reflect the internal structure and characteristics of images, and also have less sparsity [7]. Overcomplete model more in line with the mechanism of visual feature extraction, and has a good sparse approximation performance [8, 9]. However, the asymmetry of the input space and encode space

---

P. Lu (✉) · S. Huang · C. Liu · D. Yuan · Y. Lou  
School of Electrical Engineering, Zhengzhou University, No 100 Science Road,  
Zhengzhou, China  
e-mail: lupeng@zzu.edu.cn

increases the difficulty of the sparse decomposition and the model solution [10, 11].

For the above problems, we use the energy-based models method for solving the overcomplete model, and use the response coefficient matrix instead of the base function matrix for expressing visual features to solve difficulties of the sparse decomposition and the model solution.

## 4.2 Overcomplete Sparse Coding Model

The sparse coding model is:

$$I = \sum_{i=1}^m A_i S_i + N \quad (4.1)$$

where  $I$  is a  $n$  dimensional natural image,  $A_i$  is a basis function with  $n$  dimensional vector,  $N$  is a Gaussian noise,  $s_i$  is the response coefficient,  $m$  is the number of basis functions. If  $m = n$ , formula (4.1) is a sparse coding model of complete basis, if  $m > n$ ,  $s$  is a redundant matrix, then formula (4.1) is transformed into overcomplete sparse coding model.

We assume that  $W$  is receptive field,  $A = W^{-1}$  in condition of the model of complete basis. However,  $A$  is a redundant matrix in case of the model of overcomplete basis, so it is very difficult to solve  $A$ .

To solve the above problems, we use the logarithm of probability density function to define the energy-based models, as following formula (4.2):

$$\log p(x) = \sum_{k=1}^m \alpha_k G(w_k^T x) + Z(w_1, \dots, w_n, \alpha_1, \dots, \alpha_n) \quad (4.2)$$

where  $x$  is a single sample data,  $n$  is the dimension of sample data,  $m$  is the number of receptive fields, the vector  $w_k = (w_{k1}, \dots, w_{kn})$  is constrained to the unit norm,  $Z$  is the normalization constant of  $w_i$  and  $\alpha_i$ ,  $G$  is the metric function of the sparsity of neurons response  $s$ , and  $\alpha_i$  are estimated following with  $w_i$ .

In overcomplete basis case, solving the normalization constant  $Z$  is very difficult. Therefore, we adopt the score matching to estimate the receptive field. Let us introduce score function which is defined by the gradient of logarithm of probability density function:

$$\psi(x; W; \alpha_1, \dots, \alpha_m) = \nabla_x \log p(x; w) = \sum_{k=1}^m \alpha_k w_k g(w_k^T x) \quad (4.3)$$

where  $g$  is the first-order partial derivative of  $G$ .

We used the distance square of score function between parameter model and sample data to get the objective function:

$$\begin{aligned} \tilde{J} = & \sum_{k=1}^m \alpha_k \frac{1}{T} \sum_{t=1}^T g'(w_k^T x(t)) \\ & + \frac{1}{2} \sum_{j,k=1}^m \alpha_j \alpha_k w_j^T w_k \frac{1}{T} \sum_{t=1}^T g(w_k^T x(t)) g(w_j^T x(t)) \end{aligned} \quad (4.4)$$

where  $x(1), x(2), \dots, x(T)$  are  $T$  samples.

By the above analysis, the solution process of the receptive field can be summarized as follows: looking for  $W$  to promote the objective function to minimize.

We used the gradient descent algorithm to make the objective function minimization:

$$W(t+1) = W(t) - \eta(t) \frac{\partial \tilde{J}}{\partial W} \quad (4.5)$$

where  $\eta(t)$  is the learning rate, which changes with time or iteration times.

The algorithm 1 is the learning process of overcomplete set  $W$ .

Algorithm 1: Learning of overcomplete set algorithm

Input: Sample images

Output: Overcomplete set  $W$

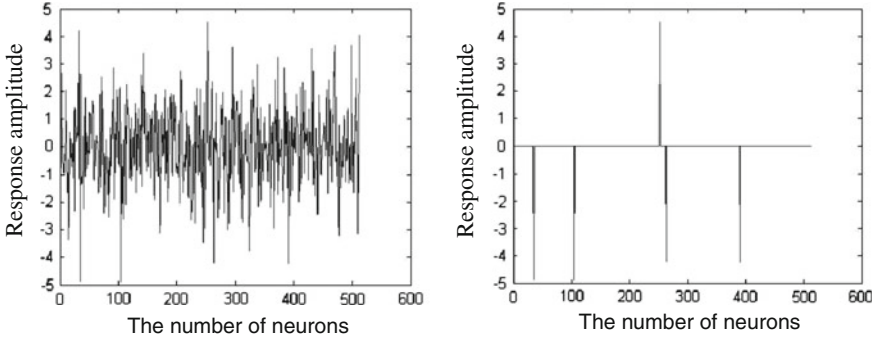
Steps:

1. Random sampling to the sample images for obtains the training samples;
2. Whiten the samples by the principal component analysis (PCA) method, and project them into whitenization space;
3. Selected the initial vector  $W_s$ , and initialize it to the unit vector, set the error threshold  $\varepsilon$ ;
4. Update  $W$  according to the formula (4.5), and normalize the unit vector, meanwhile update parameter  $\alpha$ ;
5. If  $norm(\Delta W) \leq \varepsilon$ , stop iteration, otherwise, return to step 4;
6. Stop learning, project the learning result  $W_s$  back into the original image space, then get the overcomplete set  $W$ .

### 4.3 Target Tracking Algorithm Based on the Visual Perception

Based on visual sparse and competitive response characteristics, only a small amount of neurons is activated to portray the internal structure of images and priori properties [12, 13]. We selected  $N$  neurons which have larger response as the visual feature representation of images as shown in Fig. 4.1.

We assume the difference of neurons responds between video sequence image and background image is as follows:



**Fig. 4.1** Feature extraction of image visual. **a** The response of neurons caused by image. **b** The representation of visual feature

$$h = |s_{vi} - s_{gi}| \quad (4.6)$$

where  $s_{vi}$  is the response of  $i$ th video sequence image patch, where  $s_{gi}$  is the response of  $i$ th background image patch.

The dynamic threshold is as follows:

$$\delta = \frac{1}{n} \sum_{i=1}^n |s_{vi} - s_{gi}| \quad (4.7)$$

The target tracking algorithm (TTA) is as follows:

Algorithm 2: Target tracking algorithm

Input: Video sequence image and background image

Output: The results of moving target tracking

Steps:

1. Sequential sampling to the video sequence image and background image;
2. Whiten the samples by the principal component analysis (PCA) method;
3. Calculate the neural responses of the video sequence image and background image with the formula  $s = Wx$ , and take the same number of  $N$  largest nerve responses;
4. Calculate the difference  $h$  of the neural responses of video sequence image patches and background image patches in the same location, and compared it with the dynamic threshold  $\delta$ , if  $h > \delta$ , output the results of the perception, otherwise, no further treatment;
5. Display the recognition results of the target;
6. Then enter the following frame of video sequence, return to step 1.

Flow chart of TTA is shown in the Fig. 4.2.

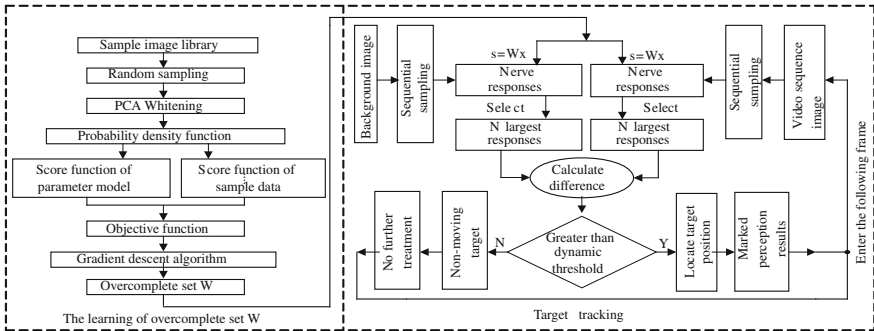


Fig. 4.2 The flow chart of TTA

## 4.4 Experiment

### 4.4.1 Learning of Overcomplete Set

Experimental environment: software system-matlab7.0, operating system-Windows XP, CPU-1.86 GHz, memory-1 GB, image resolution-512\*512.

Experimental process: Firstly, we select 10 video sequence images and use the 16\*16 sliding space sub windows for sampling each image randomly, then we get 5000 16\*16 pixels sampling patchess from one image, and 256\*50000 sampling data sets from 10 images, and then preprocess the sampling data sets, which is using the PCA method to centralize and whiten the images, and reduce the dimension to 128. The data sets of 128\*50000 is dedicated to the input of overcomplete set training. Finally, a overcomplete set representation with 512 receptive fields is estimated based on the energy-based models and the result is shown in Fig. 4.3.

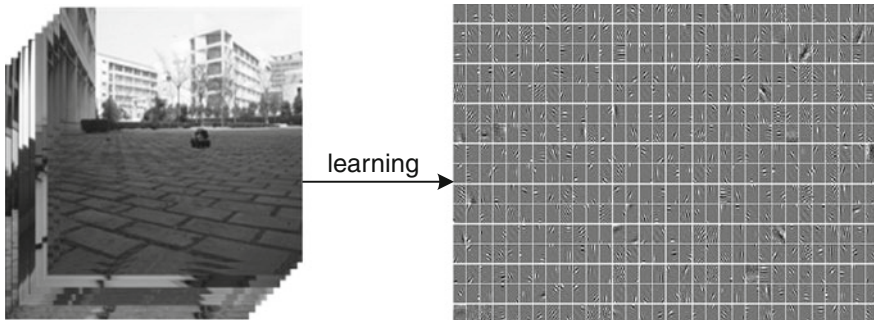


Fig. 4.3 The learning of overcomplete set

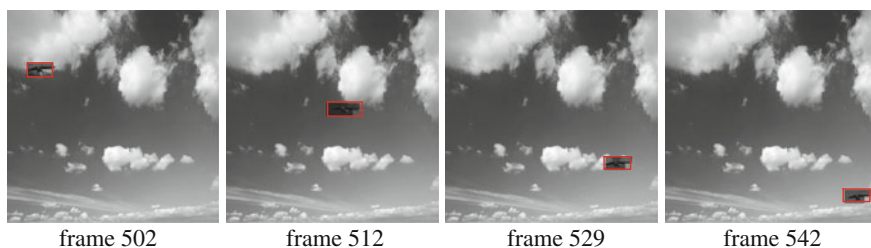


## 4.4.2 Target Tracking

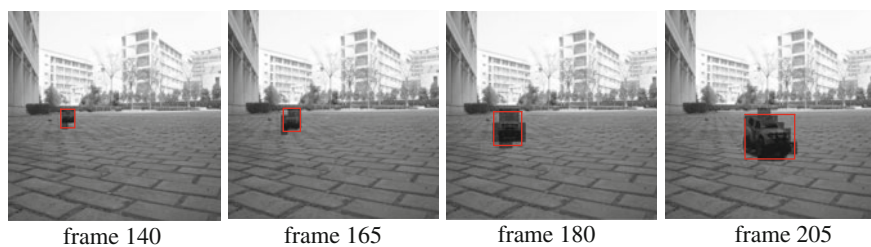
From left to right and top to bottom, we use the  $16 \times 16$  sliding space sub windows for sampling each image, and get 1024 pixels sampling patches from one image.

We designed experiments for simple background, target scale change, partial block and complete block. Results of tracking are shown in Figs. 4.4, 4.5, 4.6, and 4.7.

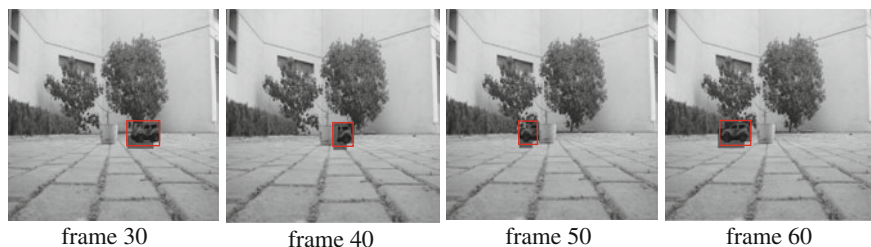
Figure 4.5, the scale and shape of target were changing in the vision. Figures 4.6 and 4.7, the target just passed behind different and similar objects in condition of the partial and complete block, so inter-class change occurs in tracking process.



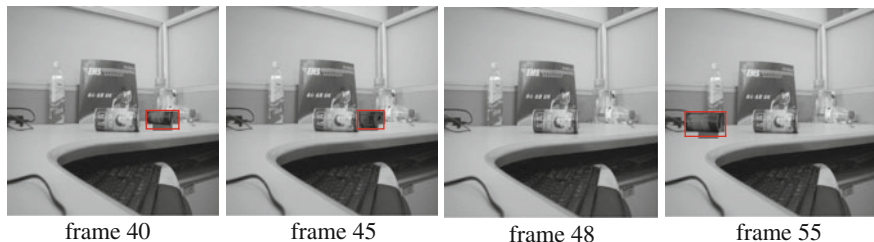
**Fig. 4.4** Tracking result of the simple background



**Fig. 4.5** Tracking result of the target scale change



**Fig. 4.6** Tracking result of the partial block



**Fig. 4.7** Tracking result of the complete block

**Table 4.1** The statistics results of three algorithms

	Video sequences (frame)	Right tracking (frame)	False discover target (frame)	False judge non-target (frame)	Wrong tracking (frame)	Recognition (%)
TTA	898	840	26	32	58	93.5
SIFT	898	663	124	111	235	73.8
Camshift	898	564	159	175	334	62.8

**Table 4.2** Time-consume comparison of three algorithms

	Video sequence (frame)	Time-consume of one frame (ms)	Total time (ms)
TTA	898	31.2	28017.6
SIFT	898	53.7	48222.6
Camshift	898	18.3	16343.4

In order to verify the validity of TTA, we compared with the typical SIFT and Camshift on the robustness, accuracy and real-time.

### 4.4.3 Analysis of Results

As can be seen in Figs. 4.4, 4.5, 4.6, and 4.7, TTA which was based on visual perception mechanism achieved tracking of target stably in condition of the block and target scale change. In the Table 4.1, error tracking frames include the false discovery and false judge non-target: the false alarm and missed alarm, the TTA algorithm improves the accuracy of target tracking compared with SIFT and Camshift. It can be seen from the Table 4.2, the time-consume of TTA algorithm is less than the SIFT, and more than the classic Camshift slightly, but to meet the real-time requirement.

## 4.5 Conclusion

By simulating visual perception mechanism, we established a new kind of target tracking algorithm TTA, and its accuracy and robustness have been improved. TTA algorithm achieved tracking of target stably when occurred scale change of target and block interference, and also target deformation and inter-class exchange at the same time. The furthermore work is we will take further research combined with high-level visual semantics, such as attention and learning mechanism.

**Acknowledgments** The work for this paper was financially supported by the National Natural Science Foundation of China (NSFC, Grant No: 60841004, 60971110, and 61172152).

## References

1. Meng LF, Kerekes J (2012) Object tracking using high resolution satellite imagery. *IEEE J Sel Top Appl Earth Obs Remote Sens* 5(1):146–152
2. Yin MH, Zhang J, Sun HG, Gu WX (2011) Multi-cue-based CamShift guided particle filter tracking. *Expert Syst Appl* 38(5):6313–6318
3. Wang ZW, Yang XK, Xu Y, Yu SY (2009) CamShift guided particle filter for visual tracking. *Pattern Recogn Lett* 30(4):407–413
4. Yao MH, Zhu H, Gu QL, Zhu LC, Qu XY (2011) SIFT-based algorithm for object matching and identification. *Remote Sens Environ Transp Eng* 271:5317–5320
5. Yu CB, Zhang J, Liu YX, Yu T (2011) Object tracking in the complex environment based on SIFT. *IEEE Commun Softw Netw* 141:150–153
6. Koldovský Z, Tichavský P (2011) Time-domain blind separation of audio sources on the basis of a complete ICA decomposition of an observation space. *IEEE Trans Audio Speech Lang Process* 19(2):406–416
7. Casaletti M, Maci S, Vecchi G (2011) A complete set of linear-phase basis functions for scatterers with flat faces and for planar apertures. *IEEE Trans Antennas Propag* 59(2):563–573
8. Mohimani H, Babaie-Zadeh M, Jutten C (2009) A fast approach for overcomplete sparse decomposition based on smoothed  $\ell^0$  norm. *IEEE Trans Signal Process* 57(1):289–301
9. Labusch K, Barth E, Martinetz T (2009) Sparse coding neural gas: learning of overcomplete data representations. *Neurocomputing* 72(7–9):1547–1555
10. He ZS, Xie SL, Zhang LQ, Andrzej C (2008) A note on Lewicki-Sejnowski gradient for learning overcomplete representations. *Neural Comput* 20(3):636–643
11. Hyvarinen A, Hurri J, Hoyer PO (2009) *Natural image statistics*. Springer, Berlin, pp 289–444
12. Sun H, Sun X, Wang HQ, Li Y, Li XJ (2012) Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model. *IEEE Geosci Remote Sens Lett* 9(1):109–113
13. Dai DX, Yang W (2011) Satellite image classification via two-layer sparse coding with biased image representation. *IEEE Geosci Remote Sens Lett* 8(1):173–176

# Chapter 5

## Positioning Tropical Cyclone Center in a Single Satellite Image Using Vector Field Analysis

Jinfeng Yang and Haoren Wang

**Abstract** Tropical cyclone (TC) center location or eye fix is the first and inevitable procedure in TC forecasting, but mostly done manually in practice. Here we present a novel and objective method to locate TC center by compensating conventional pattern matching and vector field analysis methods together. The experimental results show satisfactory results in comparison with the best track data from the China Meteorological Administration especially when the spiral patterns of TC are well structured. Our method being introduced, TC forecasters can fix TC eye objectively and easily from data of all kinds of meteorological satellites.

**Keywords** Typhoon · Location · Ridge · Gradient field · Equiangular · Spiral

### 5.1 Introduction

A *tropical cyclone (TC)* is a low-pressure system with a warm core that forms over tropical and subtropical waters [1]. In TC forecasting, it is the most basic problem that accurate and timely TC center located from all kinds of remote sensed data. Typically, a majority of attempts to fix TC eyefall into two categories: *pattern matching (PM)* and *vector field analysis (VFA)* [2].

Briefly speaking, PM is to define a model for TC with and eye and then match it with image data in computer under some criterion. Different PM approaches mainly vary in different models, such as [3–5]. Though intelligent searching algorithms can be introduced to low calculating time [6], time-honored pattern matching method is becoming time-consuming as the models are getting more complicated and the spatial resolution increasing.

---

J. Yang (✉) · H. Wang  
Tianjin Key Lab for Advanced Signal Processing, Civil Aviation University of China,  
Tianjin, P.O.Box 9 People's Republic of China  
e-mail: jfyang@cauc.edu.cn

In VFA, the first step is to construct vector field from consecutive images [2]. Next step is to analyze the vector field to find a TC center, including *algebraic analysis* (AA) and *structural analysis*. The idea of AA is to decompose a vector field into a sum of solenoidal and irrotational components [7]. Singular point or critical point can be considered as the eye of TC, such as in [8, 9]. Structural analysis, proposed in [2, 10], is a geometric analysis method. Since the vectors in field are rotated and expanded to sectors which seemed like flashlights and the method employs a voting schema, we would like to name this unique algorithm as *rotated flashlight voting* (RFV). It takes use of the geometric feature of spiral that the angle between radial and tangent lines of log spiral is always constant. RFV is fast and performs well in noise tolerance. However, there is an assumption that in the step of vector field construction that the movement of TC is rigid motion. In this scenario, RFV can show performance only when data with high temporal resolution are available.

To the best of our knowledge, we find that both PR and VFA methods have their advantages and disadvantages. If they can be combined together, PR and VFA would compensate each other. In this study, we developed a schema to make the two kinds of methods into one. The proposed method is capable of locating TC center from any data source without taking temporal resolution into account, which is convenient for forecasters to fix the TC eye fast, accurately, and objectively.

## 5.2 Methodology

### 5.2.1 About Logarithmic Spiral Model

In this sub-section, we described the features of logarithmic spiral—the model of our schema. *Logarithmic Spiral* is defined in polar coordinate as:

$$\rho(\theta) = a.e^{b\theta} = a.e^{\theta \cot \alpha} \quad (5.1)$$

where  $a$ ,  $b$  and  $\alpha$  are TC-specified parameters. As can be seen in Fig. 5.1, the value of  $\beta$  equals to:

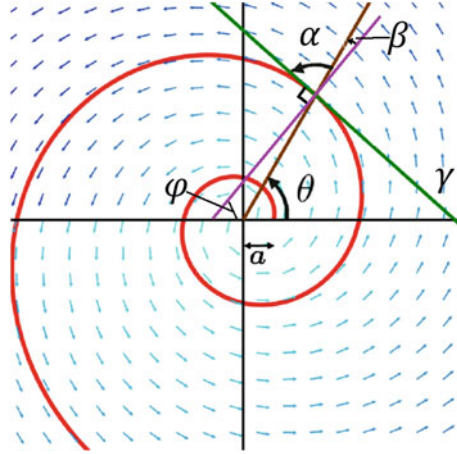
$$\beta = \theta - \varphi = \frac{\pi}{2} - \alpha. \quad (5.2)$$

There is a feature of log spiral that the value of  $\alpha$  in every point of the spiral is always identical. This can be proved as follows.

$$\tan \gamma = \frac{dy}{dx} = \frac{dy}{dx} \bigg/ \frac{dx}{d\theta} = \frac{\rho'(\theta) \tan \theta + \rho(\theta)}{\rho'(\theta) - \rho(\theta) \tan \theta}. \quad (5.3)$$

And, as seen in Fig. 5.1 that,  $\gamma = \alpha + \theta$ , then,

**Fig. 5.1** Parameters of a spiral flow, where  $\alpha$  is the angle between the radial line and the tangent line of the spiral,  $\beta$  is the angle between the radial line and the perpendicular of tangent line (gradient of the spiral),  $\gamma$  is the angle of tangent line, and  $\varphi$  represents the direction of gradient



$$\tan \alpha = \tan (\gamma - \theta) = \frac{\tan \gamma - \tan \theta}{1 + \tan \gamma \tan \theta}. \quad (5.4)$$

Then, substituting formula (5.3) into formula (5.4) yields

$$\tan \alpha = \frac{\rho(\theta)}{\rho'(\theta)}. \quad (5.5)$$

And then, substituting formula (5.1) into formula (5.5) yields

$$\tan \alpha = \frac{a \cdot e^{b\theta}}{ab \cdot e^{b\theta}} = \frac{1}{b}. \quad (5.6)$$

So the logarithmic spiral is also named *equiangular spiral*. Then it is obvious that  $\beta$  is also a constant, according to formula (5.2). This feature will be helpful our schema.

## 5.2.2 Steps

### 5.2.2.1 Preprocess

As we know, clouds are usually colder than the underlying sea surface, continent or other backgrounds. Colder *brightness temperatures* correspond with higher grayscales of IR images [11]. In this paper, we first extracted the cloud mass in image by threshold. In such a way, brighter pixels were retained and the rest were assigned the value zero. Then the extracted image was *cropped* for shortening time cost. We then applied *histogram equalization* on the image to enhance the contrast. The *average filtering* was applied in this step to eliminate *Gaussian noise*. After preprocessing, the noise involved was reduced while preserving the details.

### 5.2.2.2 Construct Vector Field in a Single Image

In our algorithm, we would like to construct vector field in a single image rather than a sequence of images. The spiral features give clues to fix eye of TC, whether the eye is recognizable or not. Therefore, the vector field we got shall be related to the spiral trend of cloud instead of motion trend. As we know, an image can be considered as a scalar field. The question turns out to be *how to construct vector field from a scalar field*. In *vector calculus*, the gradient of a scalar field is a vector field. The gradient is defined as:

$$\nabla f(x, y) = [G_x \ G_y]^T = \left[ \frac{\partial f}{\partial x} \ \frac{\partial f}{\partial y} \right]^T, \quad (5.7)$$

$$\varphi(x, y) = \tan^{-1}(G_y/G_x), \quad (5.8)$$

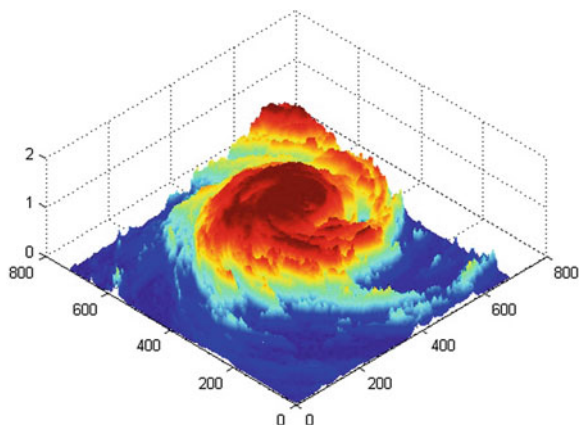
where  $f(x, y)$  represents the gray-scale image. Thus, we tried gradient operator on the image that we had previously preprocessed. And then the gradient field was averaged. However, after these procedures, the gradient field we got was full of unordered and disorganized vectors which would be helpless in TC eye fixing. So far, the gap between scalar field and vector field is still remaining. The problem is then *how to filter out the noise in the gradient field*.

Inspired by the concept in [12, 13], we found that the pixel values of 2-dimensional imagery could be viewed as z-axis values in 3-dimensional space. It is very similar to a mountain, and we call it a *cloud mountain*, illustrated in Fig. 5.2. Also, as we know, the pixel value in IR imagery is a function of temperature, which is related to the cloud height. The trend of circulating cloud looks like mountain crest which can tell the range of a mountain. To extract the *venation* of a cloud mountain, in this paper, we determined the candidate ridge points [13] by analyzing the local maxima. If a point is a ridge point, the point will be local maxima along the gradient direction formula (5.8). The contrary of this happens when it is a valley point. We selected the points that were relative maxima to the gradient direction as valid points. In this manner, the gradients which are local maxima are filtered as valid gradients. Then helpful information about the circulating trend was extracted while useless noises were abandoned at the same time. In addition, the number of vectors to be processed is reduced which makes the algorithm more efficient in time. The gap had been filled out so that a feasible vector field was obtained.

### 5.2.2.3 Analyze the Vector Field to Locate the Center

Since a vector field was obtained in the last step, we could then employ the RFV [10] (Fig. 5.3). In [10], to offset the effect of noise and rounding errors of motion field estimation, sector span  $\sigma$  and a small tolerance angle  $\sigma^+$  is introduced, and another parameter  $\acute{a}$  is brought into control the distance of influence.

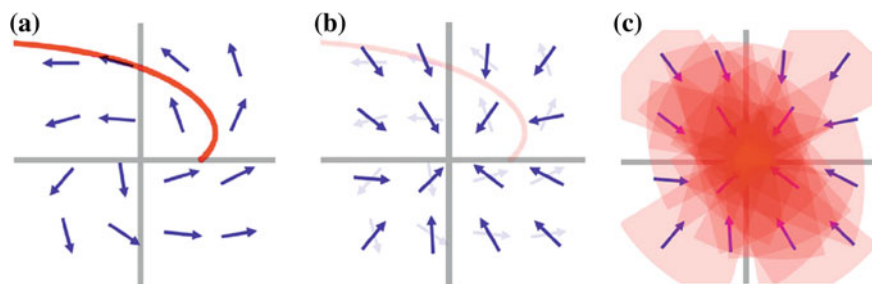
**Fig. 5.2** 3D view of a TC, the values on z-axis are the grayscales of image which are related to the cloud height. It looks like a mountain composed by cloud



Actually, the vector field we got in previous procedure is a gradient field. Moreover, the vector field used in the RFV algorithm [10] is a directional field. Since the direction field is, in principle, perpendicular to the gradients [13], here is a question that *can the rotation step in RFV algorithm be canceled*.

The RFV algorithm is based on the property of logarithmic spiral that the angle between its tangent in any point and the corresponding radial is always a constant as  $\alpha$  (see formula (5.6) and Fig. 5.1). The parameter  $\alpha$  controls the directions of vectors in the field with respect to the radials. And the optimal rotation angle is  $\pi - \alpha$  with  $\alpha$  unknown in practical situation. If the rotation step in RFV is canceled, then  $\alpha = \pi/2$ ,  $\cot \alpha = 0$ , the logarithmic spiral will degenerate to a circle. However, it is obvious that not every TC spiral feature is circular. Therefore,  $\alpha = \pi/2$  is not enough, which means the gradients need to be rotated more than  $90^\circ$ . In this paper, different from the solution in work [10], we tackled this problem with another procedure that we canceled the rotation at first and verified the candidate centers at last.

As mentioned in [2], for TC eye fix in northern hemisphere, typically varies from  $100^\circ$  to  $102^\circ$ . Therefore, in [10], the optimal  $\alpha$  is set to  $101^\circ$ , with  $\sigma^+$  of at



**Fig. 5.3** The rotated flashlight voting (RFV) algorithm, **a** construct vector field first, **b** rotate every vector at a same angle, **c** expand the rotated vectors to sectors to vote for the center



least  $2^\circ$  to make sure sectors from fields in all the typical cases are wide enough to cover the center. This solution is sound in typical spiral cases while some TCs have symmetrical circular structure in maturity period [10]. In our method, instead of rotating the gradient vectors, we enlarged the span of sectors by increasing the value of  $\sigma$  to make sure both typical and untypical TC centers were covered. And parameter  $\sigma^+$  is canceled for simplification. The real center should be among the most covered points or the most voted ones. Thus, we determined that the pixels whose numbers of voting tickets were among the top  $1/m$  in all,  $m$  is a human-specified parameter, as the candidate centers. Note that, in all candidate pixels, if a pixel is the real center of TC, the angle between every valid gradient and the corresponding radial should be totally identical (see formula (5.2) and Fig. 5.1). In practice, these angles should be approximately the same. In other words, the sum of the variances of angle  $\beta$  should be the smallest (see formula (5.2) and Fig. 5.1). The real center is defined as:

$$C_{real} = \underbrace{\arg \min}_c \left\{ \sum_i Var(|\theta_i^c - \varphi_i|) \right\}, \quad (5.9)$$

where  $C_{real}$  is the position of the real center,  $\{c\}$  represents the set of candidate centers,  $i$  depicts every pixel of valid gradients,  $\theta_i^c$  means the polar angle from candidate center  $c$  to pixel  $i$ , and  $\varphi_i$  means the angle of valid gradient in position  $i$ .

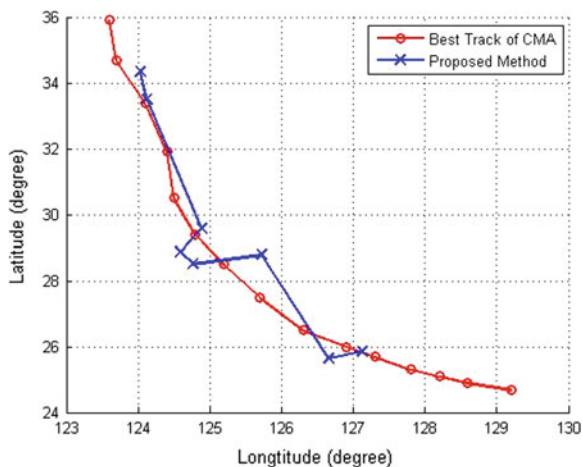
### 5.3 Experimental Results

The data used in our experiment was IR-images derived from FengYun-3A and FengYun-3B with 1 km spatial resolution. Due to higher spatial resolution, images from polar-orbiting satellites are favorable for our scheme. In our experiments, 8 IR-images related to TC Muifa (August 5–7, 2011) were chosen for testing the proposed algorithm, see Table 5.1. The best track data of the CMA (China Meteorological Administration) is linearly interpolated to evaluate the proposed locating method.

Since our data was low in temporal resolution, the track in Fig. 5.4 is a little discontinuous and unsmooth. However, it is shown that the track of our method is in general similar with the observed track. The mean absolute error for our examples was found as  $0.28^\circ$  in latitude and  $0.24^\circ$  in longitude respectively. It must be pointed out that the presented method is theoretically well performed in condition that bare bones of TC is equality in pitch. Thus, the accuracy of the presented algorithm is mainly affected by the symmetry of the TC. In other words, it depends on how well structured the TC is. Another factor relating to accuracy are the value of parameter  $\sigma$  in RFV [10]. In our work,  $\sigma$  was set to  $\frac{\pi}{4}$ , which could be optimized by some hit and trial experiments in future work.

**Table 5.1** Experimental result for verifying our proposed algorithm

Universal time	Satellite brand	CMA Lat (°)	CMA Lon (°)	Our Lat (°)	Our Lon (°)	Error Lat (°)	Error Lon (°)
08050550	FY3A	25.69	127.31	25.86	127.12	0.17	-0.19
08051235	FY3B	26.05	126.84	25.65	126.66	-0.4	-0.18
08060530	FY3B	28.42	125.24	28.79	125.72	0.37	0.48
08060840	FY3A	28.90	125.02	28.52	124.77	-0.38	-0.25
08061020	FY3A	29.15	124.91	28.87	124.59	-0.28	-0.32
08061220	FY3B	29.46	124.78	29.58	124.89	0.12	0.11
08070820	FY3A	33.91	123.94	33.49	124.12	-0.42	0.18
08071000	FY3A	34.27	123.83	34.33	124.03	0.06	0.20

**Fig. 5.4** Comparison of our proposed method and the best track data from CMA

## 5.4 Conclusion

In summary, we proposed a novel algorithm for TC center locating by making the pattern matching and vector field analysis methods combined together and compensated each other. By the concept of cloud mountain, we obtained valid gradient vectors that are also ridge or valley points from a single satellite IR image which is a scalar field, so that gradient field related to the mountain trend would be feasible for analyzing TC center. And we verified the candidate centers afterwards according to the equal pitch property of log-spiral and circle instead of rotating the gradient vectors more at first, so that the robustness of our algorithm to different TC life period is enhanced without so much extra calculation. In this way, the proposed method benefits from both sides and can be used in images from solar orbiting without taking temporal resolution into account. However, one of the limitations of this technique is its inability to handle the case where the field is viewed orthographically but not directly from the top [10]. And further work will

be required to test more TCs of different structure and different period of life cycle before operational application.

## References

1. Roy C, Kovordányi R (2012) Tropical cyclone track forecasting techniques—a review. *Atmos Res* 104–105:40–69
2. Wong KY, Yip CL, Li PW (2007) A novel algorithm for automatic tropical cyclone eye fix using Doppler radar data. *Meteorol Appl* 14:49–59
3. Yan WK, Lap YC, Ping Wah LI et al (2004) Automatic template matching method for tropical cyclone eye fix. In: Proceedings of the 17th international conference on pattern recognition, ICPR 2004. Institute of Electrical and Electronics Engineers Inc., Cambridge, pp 650–653
4. Yurchak BS (2007) Description of cloud-rain bands in a tropical cyclone by a hyperbolic-logarithmic spiral. *Russ Meteorol Hydrol* 32:8–18
5. Wei K, Jing Z-L (2010) Spiral band model optimization by chaos immune evolutionary algorithm for locating tropical cyclones. *Atmos Res* 97:266–277
6. Wei K, Jing Z-L, Li Y-X et al (2011) Spiral band model for locating tropical cyclone centers. *Pattern Recogn Lett* 32:761–770
7. Wong KY, Yip CL (2007) A fast and noise-tolerant method for positioning centers of spiraling and circulating vector fields. In: 8th Asian conference on computer vision, ACCV 2007, November 18–22, 2007. Springer, Tokyo, pp 764–773
8. Chiao-Fe S, Jain RC (1994) Vector field analysis for oriented patterns. *Pattern Anal Mach Intell IEEE Trans* 16:946–950
9. Wei L, Ribeiro E (2012) Detecting singular patterns in 2D vector fields using weighted Laurent polynomial. *Pattern Recogn* 45:3912–3925
10. Wong KY, Yip CL (2009) Identifying centers of circulating and spiraling vector field patterns and its applications. *Pattern Recogn* 42:1371–1387
11. Bai Q, Wei K (2011) Cloud system extraction in tropical cyclones by mountain-climbing. *Atmos Res* 101:611–620
12. Jaiswal N, Kishtawal CM (2011) Automatic determination of center of tropical cyclone in satellite-generated IR images. *IEEE Geosci Remote Sens Lett* 8:460–463
13. Bazen AM, Gerez SH (2002) Systematic methods for the computation of the directional fields and singular points of fingerprints. *IEEE Trans Pattern Anal Mach Intell* 24:905–919

# Chapter 6

## Target Extraction Study on the Vision System of Apple Picking Robot

XinWen Cheng and XueQiang Shi

**Abstract** This paper studied the color images of mature apple in natural environment, there is a general distinction between mature apple's color and the background. Apple and background distributed in different area of color space. According to this characteristic, this paper proposed an object extraction algorithm based on sample color space. First we construct the sample color space in  $L^*a^*b^*$  space by using apple samples' image and using mathematical morphology to optimized it. Then recognised the apple target according to the sample color space. For the small target of depth of field and serious keep out targets, we make them into binarized images and use morphology structure element again to processing them. At last we got the ideal segmentation effect with high recognition rate.

**Keywords** Image segmentation · Sample color space · Binarization · Mathematical morphology

### 6.1 Introduction

China is the world's largest producer of apple and occupied an important position of fruit production in the world. However, our apple picking is still artificial completed, efficiency is very low. Using computer vision to detect and recognize apple in natural scene of image can realize its automatic picking. Image

---

X. Cheng (✉)

Jiang Cheng College, China University of Geosciences, 4th Floor, Administrative Building  
No 8 XiongTing Bi Road, ZhiFang, JiangXia, Wuhan, China  
e-mail: chxw377@126.com

X. Shi

Faculty of Information Engineering, China University of Geosciences, 3rd Floor, No 388  
LuMo Road, HongShang, Wuhan, China  
e-mail: 574980758@qq.com

segmentation is one of the crucial steps. So this paper studies the apple image's segmentation of natural conditions and provides fast complete object segmentation method for subsequent robot visual identification and automatic picking. At present, there are many people study on the color image segmentation at home and abroad. In traditional image target recognition, image will often directly make into binary and extract the object. For example Niu xi-quan etc., [1] make the color mature apple image under natural scene into gray image, then segment into binary images through threshold. Finally using mathematical morphology filter to remove high frequency noise. That segmentation method fully avoid human intervention and fast.

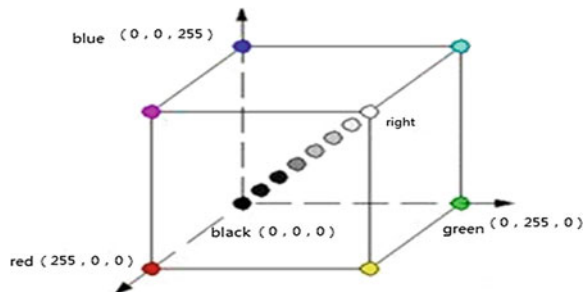
However the use of gray image lost a lot of useful image information. For the apple image with complex background, it cannot achieve the ideal effect with directly binarization. Using multi-threshold segmentation was easy to cause target and background segmenting into a single tablet that can not recognise of target [2]. Clustering algorithm and stochastic model algorithm are difficult to meet the real-time effects because of the complex computation and cost of processing time [3].

According to the characteristic that the apple target have obvious difference with the background apple on color, we carry on the research from color space. First we establish a apple sample color space and then segment the apple image according to the sample space. For the small target of depth of field and serious keep out targets, we use morphology to process recognise them and got the ideal segmentation result.

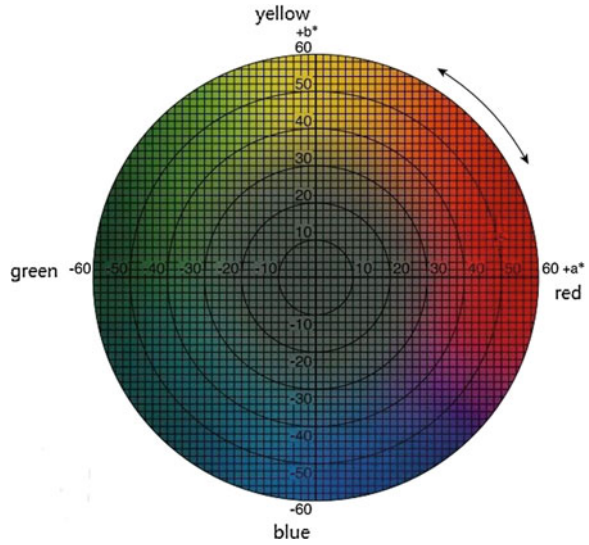
## 6.2 Color Space System

In the design principle of chromatology, there are some common color spaces like RGB, HSV, YCb,  $L^*a^*b^*$  and so on. In this paper we mainly introduced the RGB and  $L^*a^*b^*$  color space involved first. RGB color space is established according to the low of trichromatic theory and one of the basic color spaces. It corresponding a cube of cartesian coordinate system. RGB represent the three axes respectively. As shown in Fig. 6.1: When RGB took 0 is origin of coordinates is black, all take the maximum means white.

**Fig. 6.1** RGB color space diagram



**Fig. 6.2** L\*a\*b\* color space a\*b\* diagram



L\*a\*b\* model is one of the two color spaces recommended by the international lighting committee in order to uniform color difference evaluation method and its advantage is luminance component separation. We can use the space distance to present the differences of two colors. In L\*a\*b\* space, L\* means luminance, its value is 0–100. a\* present red and green and their change. The positive value of a\* means red and negative means green, ranging from –60 to 60. b\* means the change from yellow to blue, positive of b\* means yellow and negative means blue. Ranging from –60 to 60. Show as Fig. 6.2. The transformation formulas from RGB space to L\*a\*b\* space is [4]:

$$L^* = 116(0.299R + 0.587G + 0.114B)^{1/3} - 16 \tag{6.1}$$

$$a^* = 500[1.006(0.607R + 0.174G + 0.201B)^{1/3} - (0.299R + 0.587G + 0.114B)^{1/3}] \tag{6.2}$$

$$b^* = 200[(0.299R + 0.587G + 0.114B)^{1/3} - 0.846(0.066G + 1.117B)^{1/3}] \tag{6.3}$$

In L\*a\*b\* color space, a\* and b\* have nothing to do with the brightness of the component. We can weaken the radial and shade’s influence of segmentation results in natural environment and use the two component to build apple sample color space.

## 6.3 Building Apple Sample Color Space

### 6.3.1 Sample Collection

Taking many mature apple pictures from apple orchard with a digital camera, cutting out mature apple areas

Taking the same variety of mature apple pictures bought back from market in all sorts of natural light and background conditions like early morning, noon, evening, sunny and cloudy with ground, leaves or weeds.

### 6.3.2 Building Sample Space

Building an  $120 \times 120$  two-dimensional zero array  $A[120][120]$ , then transforming the apple sample color image collected from RGB color space to  $L^*a^*b^*$  space. Using array  $A$  to count the number pixel  $(i, j)$  of sample appeared in space  $a^*b^*$  ( $L^*$  is ignored). If pixel  $(i, j)$  appears one time, then  $A(i, j)$  corresponding increases one. At last we can got the 2D color distribution density array of sample. Selecting an appropriate threshold value depending on how many the samples are. Probability value which pixels appears less than threshold will be assigned to zero. Finally the pixel set in  $a^*b^*$  space corresponding by the non zero array is the apple sample color space. In order to reduce the random error and noise on the influence of sample color space and improve the recognition accuracy, we optimized the color space with the morphology: The statistical array can take as a two dimensional image after threshold processing, Zero remain unchanged, nonzero namely assignment is 1, Two-dimensional array becomes a binary map. Mathematical morphology open operation and closed operation are made on the binary image, we can eliminate holes, saws and small gaps so as to reduce the error noise, accidental pixels and edge transition pixels' impact on the results of the sample space. Finally the pixel sets in  $a^*b^*$  space corresponded by the binary image is the apple sample color space.

## 6.4 Mathematical Morphology

The mentioned mathematical morphology above introduced as follows:

The most basic operations of mathematical morphology are corrosion and inflation [5]. Corrosion: supposing  $P$  is input elements,  $Q$  is structure elements.  $Q$  correds  $P$ , which shows  $P \odot Q$ . As follows:

$$P \odot Q = \{x : Q + x \subseteq P\} \quad (6.4)$$

If  $Q$  is regarded as a template,  $P \odot Q$  composes of all points that can be filled into the internal template of  $P$  when they are in the processing of translation template.

Inflation is the dual operation of corrosion. The follow is the definition that  $Q$  influts  $P$ :

$$P \oplus Q = \cup \{P + q : q \in Q\} \quad (6.5)$$

The Morphological opening and closing operation are different combinations of corrosion and inflation. The image  $Q$  is used for the opening operation of image  $P$ , which is represented as  $P \ominus Q$ .

$$P \ominus Q = (P \odot Q) \oplus Q = \cup \{Q + x : Q + x \subseteq P\} \quad (6.6)$$

There are two functions of the opening operation. First, the disk is applied in the opening operation so that the internal edge can be polished. In additional, the discal rounding can get the low pass filtering effect.

Closing operation is the dual operation of opening, and is defined that inflation is processed before corrosion.  $P \bullet Q$  means that  $Q$  is used in the closing operation of  $P$ .

$$P \bullet Q = (P \oplus B) \odot B \quad (6.7)$$

The closing operation can remove foraminule and concave department of images and fill in small gaps or hole of target interior. It also can lap an short intermittent to product the connectivity.

## 6.5 Apple Target Recognition Test

### 6.5.1 Experimental Design and Realization



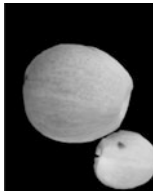
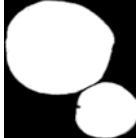


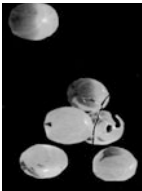
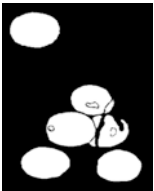


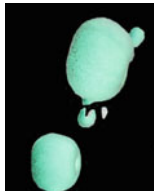
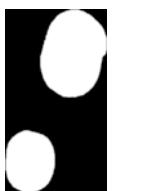
Transforming the apple color image of natural environment into a  $L^*a^*b^*$  space. Inquiring the apple color image's pixels at the sample color space constructed, the pixels are perceived as apple target's pixels if they can be found in sample space, otherwise, the background. Then we realized the apple target recognition and segmentation. In practice, we selected apple color image pixels' seven areas for statistics judgment.

Transforming the apple color image of natural environment into  $L^*a^*b^*$  space, treating  $a^*b^*$  pixels array as  $S[120][120]$ , if more than half of the pixels in seven-neighbourhood of  $S(i,j)$  can found in sample color space,  $S(i,j)$  can be treat as apple pixel, otherwise treat as the background pixel. Such processing can effectively inhibit noise influence on the result and can distinguish apple target edge pixels resultful. Programming to realize the above algorithm and experiment.

We took 500 mature apple images in the natural environment with all kinds of lighting conditions using a digital camera and constructed the apple sample color space. Then extracting apple from the image of natural environment conditions. Part of the division examples and results as is shown in Table 6.1.



Table 6.1 Experimental result

	Original image	Segmentation image	Graying and inverting	Morphology processing
Uncovered				
Partial covered				
Far and close shot				

### 6.5.2 Analysis

From the above a series of processing results we can see that, the algorithm used in this paper can achieve full recognition of the same depth of field and uncovered goals. It can also inhibit the influence of the light on the result. The vision of far depth of field cannot be completely identify and differentiate because of that color is the sole identify factor in sample color space. In the practical production, the far apple targets cannot be picked by the picking robot in one time and they are small areas in pictures. So we treated the far apples as background. We use morphology on the basis of color sample space algorithm to process the images with small target. We choose a quarter of apple's radius in close shot as the structural element's radius to remove the small target at the same time do not affecting the apple target recognition with close shot. Show as Table 6.1, row four, column five. After morphology processing we also know target can be recognized if the apple targets are partial covered by leaves and petioles meanwhile apples' area contain the obstruction area image. However, if contour lines of apple targets are covered by leave or branches, integrity of target recognition will be affected. Show as Table 6.1, row three, column five. This result also to provide a reference for robot design avoiding of the branches and obstacles, reducing the damage of apple tree, and improving the service life of the machine.

## 6.6 Result

The algorithm based on the apple sample color space proposed and optimized with morphology in this paper is suitable for apple's segmentation when the mature apple and background have obvious difference. The segmentation effect is remarkable. This algorithm can also inhibit the affect of light to the result.

When the contour line of the apple was covered by the branches, the apple target can be incompletely recognized. This paper also treat the deep small target objects as background, such processing considered the actual production situation and also increased the efficiency of target segmentation.

For the apple target can not complete recognized with partial covered, in production we can use different positions of photography and multiple picking method. The algorithm in this paper has important meaning for production practice.

**Acknowledgments** Nature Science Foundation of Hubei Province (2011CDB353).

## References

1. Niu X, Liang Y (2007) Research on mature apples' color image segmentation methods in natural outdoors. *J Optoelectron Laser* 18(12):1453 (in Chinese)
2. Zhang T, Lin B, Gao R (2004) Object extraction for the vision system of fruit picking robot. *J China Agric Univ* 9(2):68
3. Shigehiko H, Katsunobu G, Yukitsugu I et al (2002) Robotic harvesting system for eggplants. *JARQ* 36(3):163
4. Yi J, Mao H, Wang X et al (2006) Automatic segmentation method for multi-tomato images under various growth conditions. *J Trans CSAE* 22(10):149
5. Shi X, Cheng X, Li C et al (2011) Research on apples' color image segmentation in natural environment. *J Anhui Agri Sci* 39(30):18993

# Chapter 7

## The Research on the Method of Traffic Area Dynamic Division and Optimization

Yirong Guo, Baotian Dong and Lei Wu

**Abstract** Based on the Temporal and spatial variation characteristics of traffic flow, this paper puts forward the traffic area dynamic division method and how to realize the merger and separation of nodes in the control area, as well as inter-regional automatic combination and split. First, according to the traffic relevance and traffic similarity between adjacent intersections in the traffic area, the application of clustering analysis method, the traffic area is divided into some dynamic control zones. Then, application node shrinkage method is to determine the key nodes of the control zone, and relying on the fuzzy cognitive map is to analyze the direct and indirect influence of the key nodes and the other nodes in the control zone. Finally, according to the traffic flow state changes forms: basic remains the same, crowded diffusion and crowded subsided, respectively, is corresponding to the dynamic changes of the control zones: remains unchanged, expansion combination and shrinkage decomposition, in order to achieve the merger and separation of the nodes in the control zone, as well as inter-regional automatic combination and split.

**Keywords** Traffic engineering · Dynamic division · Merger and separation · Fuzzy cognitive map · Control zone

---

Y. Guo (✉) · B. Dong  
School of Traffic and Transportation, Beijing Jiaotong University, Room 612, No.3  
Building, Xueyuan Apartment, No.18 JiaoDa East Road, Haidian, Beijing 100044, China  
e-mail: 10114189@bjtu.edu.cn

L. Wu  
School of Business Administration, Kyung Hee University, Seoul 136-701, Korea  
e-mail: wulei310@126.com

## 7.1 Introduction

At present, about the traffic control zone dynamic division, at abroad it was studied mainly from the degree of coupling, gravity, coordination coefficient [1], grading system classification technique [2], etc., while in domestic it was studied mainly from the angle of traffic flow, distance, cycle, etc. [3]. And the application of cycle is used to determine the changes of merger coefficient in SCATS system [4], in order to judge whether the control zone is merged.

In this paper, first according to traffic relevance and traffic similarity between adjacent intersections in the traffic area, it is divided into several traffic zones of strong internal relevance and similar traffic characteristics. Then, in order to make the control area adapt to the constant change of the traffic flow, according to the traffic flow state changes forms: basic remains the same, crowded diffusion and crowded subsided, respectively, is corresponding to the dynamic changes of the control zones: fine tuning, expansion combination and shrinkage decomposition, which achieve the merger and separation of the nodes in the control zone, as well as inter-regional automatic combination and split, thus can ease traffic congestion effectively.

## 7.2 Traffic Area Dynamic Division

Traffic zone is a set of nodes or lines with a certain traffic correlation degree and traffic similarity, which changes along with the time, correlation degree and similarity changes. It reflects the traffic characteristics of the spatial and temporal changes about the urban road network.

### 7.2.1 Traffic Correlation Degree

Traffic correlation degree reflects traffic correlation characteristic between nodes. Traffic correlation characteristic is represented by  $I_T$ , which refers to associated with the degree of traffic parameters between nodes. The calculation of sections relevance between adjacent intersections, using Whitson improved model is [5]:

$$I_T = \frac{1}{n-1} \left( \frac{n \cdot Q_{\max}}{\sum_{i=1}^n Q_i} - 1 \right) \frac{1}{1+t} \quad (7.1)$$

$n$  is from upstream intersection traffic flow into branch number;  $Q_i$  is from the upstream intersection traffic flow into the first  $i$  branch;  $Q_{\max}$  is from upstream intersection traffic flow into the maximum;  $t$  is the average travel time between two intersections, unit: min, equals to the intersection spacing  $D$  divided by the average vehicle speed  $V$ .

### 7.2.2 Degree of Traffic Similarity

Degree of traffic similarity reflects the traffic similarity between nodes. It is represented by  $I_C$ , which refers to the degree of traffic parameters similarity between each node. Calculation between adjacent intersection traffic similarity, can be used to represent by the cosine of the angle between the adjacent intersection of state vector.

Among them, the node state vector is available saturation  $X$ , cycle  $T$  and flow  $Q$  to express. For the node  $v_i$  and  $v_j$ , the state vector is  $X = (X_i, T_i, Q_i)$  and  $y = (X_j, T_j, Q_j)$  respectively, so the degree of traffic similarity  $I_C$  is:

$$I_C = \cos \theta = \frac{x, y}{|x|, |y|} \tag{7.2}$$

### 7.2.3 Control Area Division

Fuzzy C-Means Algorithm (FCMA) is a typical kind of Fuzzy clustering Algorithm. According to the calculation of each group membership degree of data to determine the set of data belong to a certain degree of clustering, this will be multidimensional data space of discrete data to classify [6]. The specific method of calculation is not presented in this paper because of space limitations.

Using fuzzy C-means clustering analysis method based on the values of sections relevance and the traffic similarity carries on the control area division. Traffic area as shown in Fig. 7.1 can be divided into five control zone, as shown in Fig. 7.2. The relevance between the control zones is much smaller compared to the control zone associated with the node. The node number of control zone is 1at least.

In Fig. 7.1, the arc refers to association impact strength between the nodes. One-way line represents the direction of the association impact strength is far stronger than another direction, whichever takes the greater value. Two-way line shows that between the two directions of association impact strength is consistent. No arc indicates that it is not connected between the intersections.

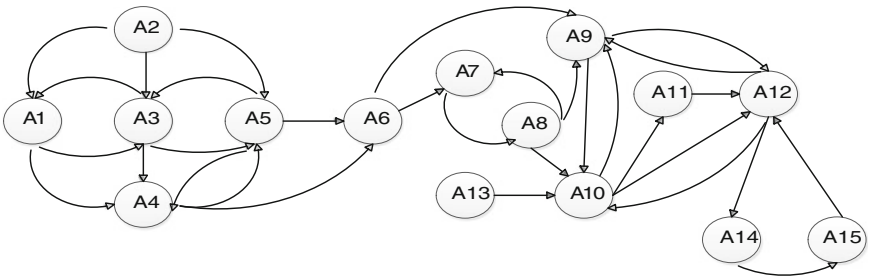


Fig. 7.1 Traffic area between nodes association diagram

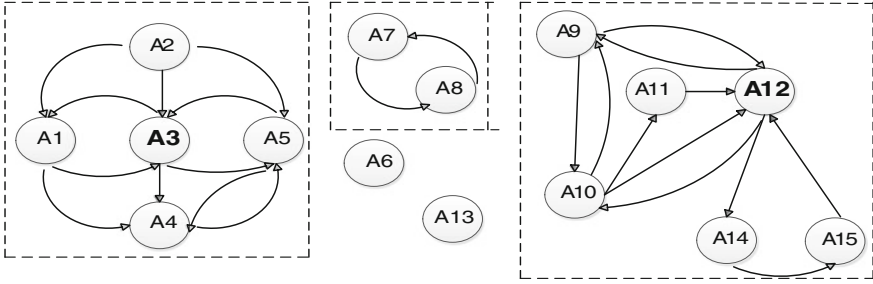


Fig. 7.2 Traffic area divided into control zones

### 7.3 Determine Key Nodes Within the Control Zone

Urban traffic network can be mapped a complex weighted network [7], namely: intersection corresponds to the node, the road corresponds to the edge, the vehicles on the road encountered impedance corresponds to the weight of the edge.

#### 7.3.1 Road Impedance Function

At present more popular and use a wide range of road impedance function is that U.S. Bureau of Public Roads is based on a large number of sections to carry out a survey and regression analysis to get BPR function model. The basic form of this model is:

$$t = t_0(1 + \alpha(q/c)^\beta) \tag{7.3}$$

$t$  is road impedance;  $Q$  is road traffic volume;  $t_0$  is free flow travel time;  $c$  is road capacity;  $\alpha, \beta$  is a regression coefficient, recommended value  $\alpha = 0.15, \beta = 4$ .

#### 7.3.2 Average Path Length

In the Complex weighted urban traffic network, the length of the shortest path between two intersection nodes  $v_i$  and  $v_j$  is  $d_{ij}$ , which is defined as the connection between the two intersections weight minimum path length. The average path length  $L$  is the average length of the shortest path between all intersections.

$$L = \sum_{i,j \in V} d_{ij} * \frac{2}{N(N-1)} \tag{7.4}$$

$N$  is the number of nodes,  $V$  is the set of nodes.

### 7.3.3 Node Important Degree

**Definition 1** Complex network node important degree is the reciprocal of the product of network node numbers and average shortest path.

The definition can get the following formula:

$$\hat{\delta} = \frac{1}{N * L} = \frac{N - 1}{2 \sum_{i,j \in V} d_{ij}} \quad (7.5)$$

$\hat{\delta}$  is node important degree, and other parameters are the same meaning as above.

Obviously,  $0 < \hat{\delta} \leq 1$ , when complex network only one node, we take the maximum value 1.

According to the weight of the road network, namely, road impedance, the use of node shrinkage method, calculates the degree of each node importance within the control zone, so as to determine the key nodes in the control zone. For each control zone in Fig. 7.2, separately calculated node importance, it can be concluded that the A3 and A12 is the key node of its control zone, while A7 and A8 node importance is equivalent, A13 node importance is 1.

## 7.4 Dynamic Adjustment and Optimization of the Control Zones

Because the network traffic flow is constantly changing, and traffic control also alters the original crowded state, control zone can't be fixed, as well as its control range is also in constant change. Therefore, we need make the appropriate adjustment according to the traffic flow change trend.

### 7.4.1 Merger and Separation of Nodes in the Control Zone

In order to determine influence degree of the direct correlation and indirect correlation between the nodes, the theory of fuzzy cognitive map is introduced. Bart Kosko proposed a fusion Zade fuzzy set concept and Alexrod cognitive map of the soft computing method on the basis of classical cognitive map, which is fuzzy cognitive map (FCM).

**Definition 2** In a control zone of the road network, the node set  $S = \{1, 2, \dots, n\}$  is the finite set, for  $S$ , there are two different nodes  $i, j$ , if the two nodes are connected and  $W_{ij} \neq 0$ , then called  $i$  to  $j$  have directly associated with the impact, denoted by  $i \rightarrow j$ .



**Definition 3** In a control zone of the road network, the node set  $S = \{1, 2, \dots, n\}$  is the finite set, for  $S$ , there are two different nodes  $i, j$ , if there are different node  $k_1, k_2, \dots, k_{(m-1)} \in S(i, j)$ , make  $i \rightarrow k_1, k_1 \rightarrow k_2, \dots, k_{(m-1)} \rightarrow j$  established, then  $i$  through the directed chain  $e_1 = \langle i, k_1 \rangle, e_2 = \langle k_1, k_2 \rangle, \dots, e_m = \langle k_{(m-1)}, j \rangle$  have indirectly associated impact, denoted by  $i \rightarrow \langle e_1, e_2, \dots, e_m \rangle \rightarrow j$ .

In order to get comprehensive associated impacts, the direct correlation classified to the indirect association will be considered together. Namely, make each node in the control zone join their associated link, and set the  $w_{ii} = 1$ , therefore,  $i \rightarrow j$  is equivalent to the link  $i \rightarrow \langle e_1, e_2 \rangle \rightarrow j$ , where  $e_1 = \langle i, i \rangle, e_2 = \langle i, j \rangle$ , via link  $\langle e_1, e_2 \rangle, i$  to  $j$  of the associated weights is:

$$W'_{ij}|_{\langle e_1, e_2 \rangle} = w_{ii}w_{ij} = w_{ij} \quad (7.6)$$

If there are  $m$  links in the road network of the control zone to make the node  $i$  associated to  $j$  ( $i \neq j$ ), the associated weights of the links are respectively:

$$W'_{ij}{}^{(1)}, W'_{ij}{}^{(2)}, \dots, W'_{ij}{}^{(m)} \quad (7.7)$$

So, comprehensive associated impact of node  $i$  to  $j$ :

$$A(i, j) = \sum_{m=1}^m W'_{ij}{}^{(m)} \quad (7.8)$$

Thus, it can not only describe the node  $i$  through the different link to  $j$  associated impact strength, but also can determine  $i$  to  $j$  of comprehensive associated impact strength. Setting the threshold value  $\lambda$  is  $i$  to  $j$  of comprehensive associated impact strength, and realize the merger and separation of the nodes in the control zone.

Reference the concept of the “Merger Index” in SCATS system, each signal cycle should be carried out once the merger index calculation. First identify the key node of control zones in each signal cycle, and calculate comprehensive associated impact strength of the key nodes and other nodes. If the associated impact strength is greater than the threshold value, then the “Merger Index” value plus 1, conversely, the “Merger Index” value minus 1. If the “merger index” is accumulated to 3, then this several intersections should be combined with coordination control; If the “merger index” is 0, then there is no need to combine together to control this several intersections, should be separated to control alone or to merge with other nodes coordination control. Adjustment process of nodes within the control zone is shown in Fig. 7.3.

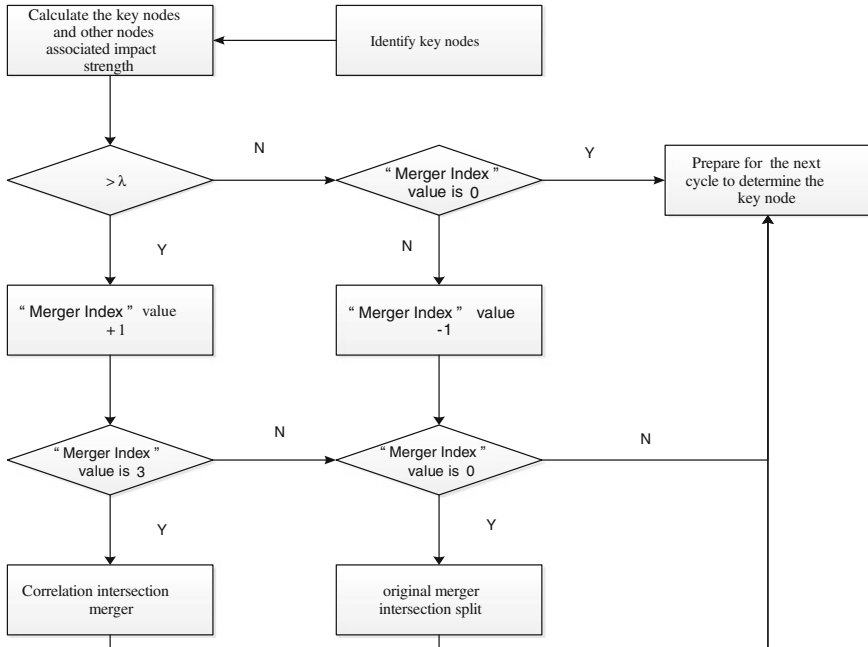


Fig. 7.3 Adjustment process chart of nodes within the control zone

### 7.4.2 Automatic Combination and Split Between the Control Zones

Automatic combination and split between the Control zones is adjusted and optimized on the basis of the original traffic area division. According to the spread of congestion and subside, for boundary intersections and sections between the control zones, by calculating traffic similarity and correlation degree of boundary intersection with adjacent intersection, the “merger index” is used to judge whether combination and split.

The boundary type of control zones (a) is shown in Fig. 7.4, for (a) provisions:

1. When two boundary intersections of the two control zone need to merge intersection B, respectively calculate traffic similarity of the intersection B and two boundary intersections to make a comparison, if the traffic similarity of control zone A’s boundary intersection is larger than intersection B’s, then B is merged into the control zone A; Conversely, B is merged into control zone C.
2. Calculating traffic similarity of the intersection B and connected intersection within the combined control zone A, if the associated impact degree is less than the threshold value and “merge index” is equal to 0, then intersection B will be separated from control zone A, considering whether automatic combination with other neighboring control zones. If the “merge index” of intersection B

and the intersections within the adjacent control zone C is also equal to 0, then intersection B will become an independent control zone, implements a single point of control.

The boundary type of control zones (b) is shown in Fig. 7.4, for (b) provisions:

1. When two boundary intersections of the two control zone need to be merged, calculated traffic correlation degree of the two boundary intersections, then respectively compared with the minimum value of the associated degrees of adjacent intersections in the control zone A and C, if the minimum value of the control zone A is much closer to the correlation degree of the two boundary intersections, then C's boundary intersection D is merged into A; conversely, A's boundary intersection is merged into C.

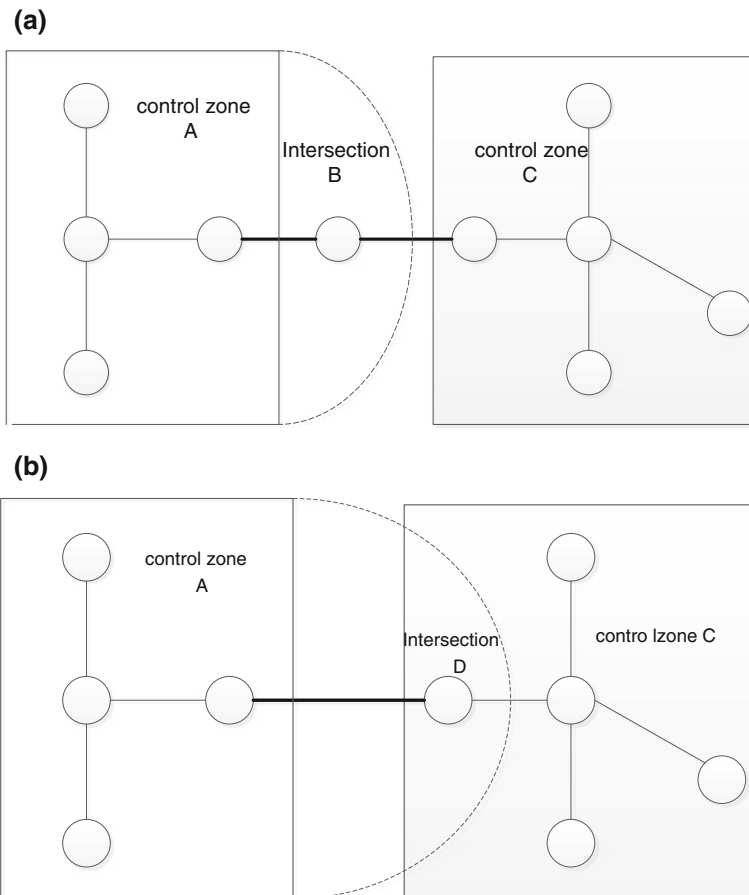


Fig. 7.4 Boundary type of control zones (a), (b)

2. Calculating traffic similarity of the intersection D and connected intersection within the combined control zone A, if the associated impact degree is less than the threshold value and “merge index” is equal to 0, then intersection D will be separated from control zone A, considering whether automatic combination with other neighboring control zones. If the intersection D is adjusted to control zone A because of congestion diffusion, so the intersection will be pulled back to the original control zone C.

So, in accordance with the rules of the merger and split, due to the crowded diffusion, original control zone can automatically extended combination, and as congestion is gradually dissipated, the extended control zone will gradually be shrunk to the original control zone.

## 7.5 Conclusion

Previous traffic zone division methods can't provide timely and effective decision-making basis for the management and control of the road network congestion. Based on the temporal and spatial variation characteristics of traffic flow, application of fuzzy theory, puts forward the traffic area dynamic division method and how to realize the merger and separation of nodes in the control area, as well as inter-regional automatic combination and split. The method is updated in real time to adjust and optimize the control zone, so as to make it adapt to the change of traffic flow, and can effectively alleviate traffic congestion.

## References

1. Hook D, Albers A (2007) Comparison of alternative methodologies to determine breakpoints in signal progression. In: The 69th annual meeting institute of transportation engineers. <http://www.trafficware.com/assets/pdfs/ooo29.pdf>
2. Zong T, Thomas U (2007) System partition technique to improve signal coordination and traffic progression. *J Transp Eng* 133(2):119–128
3. Yang QF, Chen L (2006) Division approach of traffic control work zone. *J Jilin Univ (Eng Technol Edn)* 36(2):139–142 (in Chinese)
4. Quan YS (1989) *Urban traffic control*. China Communications Press, Beijing (in Chinese)
5. Gao YF (2007) *Foundamental research on the theory of coordinated control of dynamic intersections group*. Tongji University, Shanghai (in Chinese)
6. Pu YW, Jin WD, Zhu M (2007) On cluster validity for Kernelized fuzzy c-mean algorithm. *Comput Sci* 34(2):207–210
7. Strogatz SH (2001) Exploring complex networks. *Nature* 4(10):268–276

# Chapter 8

## Multiclass Vehicle Detection Based on Learning Method

Zhiming Qian, Jiakuan Yang and Lianxin Duan

**Abstract** This paper presents a real time vehicle detection framework using learning method. This framework combines offline multiclass support vector machine and online boosting method for vehicle detection. Compare to tradition approaches, the proposed method can robust deal with multiclass vehicles and unfamiliar environment. Experiments with the city vehicle database are used to evaluate this method. The experimental results demonstrate the consistent robustness and efficiency of the proposed method.

**Keywords** Vehicle detection · Support vector machine · Online boosting

### 8.1 Introduction

The application of computer vision technology to traffic detecting and information collecting is an important subject in intelligent transportation systems, and video vehicle detection is the basic parts. Efficient and robust detection of vehicles from a video sequence has important application value in the machine vision. In recent years, tremendous efforts have been made in the field of vehicle detection to improve detection effectiveness [1–3].

In general, video vehicle detection can be divided into two categories: the learning-based approaches and the appearance-based approaches. Based on how the learning takes place over time, the learning-based approaches can be categorized as offline learning and online learning. Offline methods require all the training data to be available from the beginning of learning process. These kinds of algorithms try to produce results which are consistent with all the collected data

---

Z. Qian (✉) · J. Yang · L. Duan  
Chuxiong Normal University, Chuxiong 675000, China  
e-mail: qzhiming@126.com

samples. On the other hand, in online learning is more consistent with how the human learns and provides the machine with the ability to learn continuously and adapt all the time to its inputs. The effectiveness of a vehicle detection system employing offline training typically relies on large amount of training data. The quality and quantity of the training data certainly determine the detection capability of the system. In order to resolve the problem, the online learning methods have been an area of great recent interest in the vehicle detection.

There are some online learning methods to detect vehicle which have obtained good results [4, 5]. However, these methods have some insufficient as follows: (1) These methods can not deal with multiple class patterns. In real world, the vehicle type is various. Using single modular to detect multiclass vehicles will result in efficiency reduction. (2) Due to online learning method natural defects, such as the drift problem emerged when reusing video images to update the classifier, which also will lead to detection rate reduction.

To solve these problems, some researchers have presented many new methods. Saffari et al. [6] presented online multiclass learning method which can deal with multiple class information in the real-time environment. Grabner and Bischof [7] presented online multiple instance learning method to solve the drift problem. These methods have obtain better results in some experiments.

Offline learning strategies which means the entire training set at once which eases optimization and yields good results. In contrast, online learning strategies which means suitable for video data and environmental change. In this paper, we presented a simple and effective multiclass vehicle detection frameworks. It combines with the advantage of multiclass support vector machine (SVM) and online boosting method. The proposed method has the following characteristics: It can detect different vehicle classes in real-time environment efficiently; it just needs less time for training or learning in advance.

## 8.2 The Proposed Method

Given an input of a video sequence taken from a moving vehicle, system outputs a types and locations of the vehicles in the images. The system contains two main components: the vehicle classification and vehicle detection. In the vehicle classification, offline SVM algorithm is used to categorize classes of vehicles and create the training sample sequences for the online learning. In the vehicle detection, online learning algorithm is employed to detect the positions of the vehicles by learning the train sample set. Finally, System uses the new classifiers which are created by online learning to achieve the multiclass vehicles detection task. A general overview of the system framework can be seen in Fig. 8.1. The dashed part is completed during the initial training process of online classifiers.

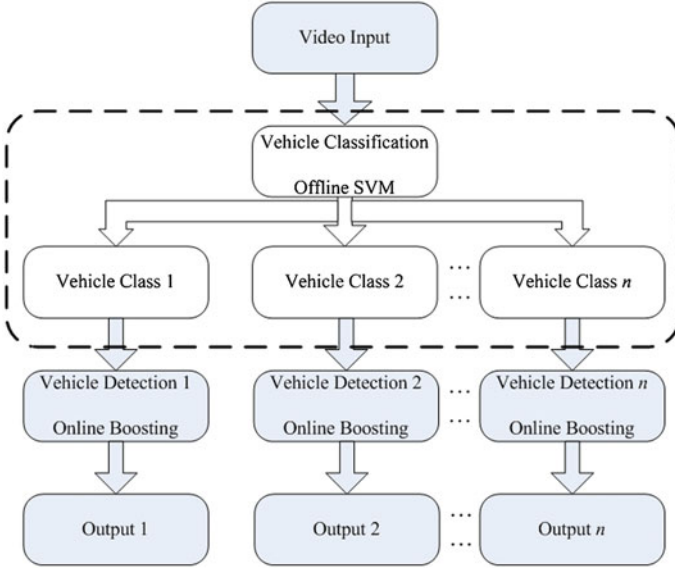


Fig. 8.1 The system framework of our method

### 8.2.1 Vehicle Classification

In vehicle classification phrase, we use multiclass SVM method in the LibSVM [8] to achieve vehicle classification. LibSVM is a useful library for SVM. It is currently one of the most widely used SVM software. LibSVM uses the one-against-one approach for multiclass learning problems [9]. If  $k$  is the number of classes, then  $k(k-1)/2$  classifiers are constructed and each one trains data from two classes. For training data from the  $i$ th and the  $j$ th classes, LibSVM solve the following two-class classification problem:

$$\min_{\omega^{ij}, b^{ij}, \zeta_t^{ij}} \frac{1}{2} (\omega^{ij})^T \omega^{ij} + C \sum_t (\zeta_t^{ij}) \quad (8.1)$$

$$\begin{cases} (\omega^{ij})^T \phi(x_t) + b^{ij} \geq 1 - \zeta_t^{ij}; \zeta_t^{ij} \geq 0, & x_t \text{ in the } i\text{th class} \\ (\omega^{ij})^T \phi(x_t) + b^{ij} \leq \zeta_t^{ij} - 1; \zeta_t^{ij} \geq 0, & x_t \text{ in the } j\text{th class} \end{cases} \quad (8.2)$$

After classifiers are constructed, use the “Max Wins” voting strategy doing the future testing: if  $\text{sign}\left((\omega^{ij})^T \phi(x) + b^{ij}\right)$  say  $x$  is in the  $i$ th class, then the vote for the  $i$ th class is added by one. Otherwise, the  $j$ th is increased by one. Then we predict  $x$  is in the class with the largest vote.

### 8.2.2 Vehicle Detection

Boosting is one of the best and thus one of the mostly applied methods in vehicle detection. Boosting for vehicle detection as described in the previous section most works offline. Hence, all training samples must be given in advance, which is not the case for video vehicle detection. Since for online learning each sample is discarded directly after an update all steps have to be online. In this paper, we use Grabner et al.'s online boosting method creating vehicle detector.

The main steps of online learning are briefly described below.

A selector  $s_n(x)$  can be considered a set of  $M$  weak classifiers  $\{h_1(x), \dots, h_M(x)\}$  that are related to a subset of features  $F_n = \{f_1, \dots, f_M(x)\}$ , where  $F$  is the full feature pool. At each time the selector  $s_n(x)$  selects the best weak hypothesis according to the estimated training error.

To start the learning process a fixed set of  $N$  selectors  $s_1, \dots, s_n$  is initialized randomly. Whenever a new training sample  $(x, y)$  arrives the selectors are updated. These updates are performed with respect to the importance weight  $\lambda$  of the current sample, which is initialized with  $\lambda = 1$ .

To update the selector  $s_n$  first all weak classifiers  $h_{n,m}(x)$  are estimated by evaluating the feature  $f_{n,m}$  on the sample image  $x$  and the corresponding errors:

$$\varepsilon_{n,m} = \frac{\lambda_{n,m}^{wrong}}{\lambda_{n,m}^{corr} + \lambda_{n,m}^{wrong}} \quad (8.3)$$

are computed. The weights  $\lambda_{n,m}^{corr}$  and  $\lambda_{n,m}^{wrong}$  are estimated from the correctly and wrongly classified examples seen so far. Then, the selector  $s_n$  selects the weak classifier  $h_{n,m^+}$  with the smallest error  $\varepsilon_n = \varepsilon_{n,m^+}$ , where  $m^+ = \operatorname{argmin}_m(\varepsilon_{n,m})$ :

$$s_n(x) = h_{n,m^+}(x) \quad (8.4)$$

According to the error  $\varepsilon_n$  the voting weight  $\alpha_n$  and the importance weight  $\lambda$  are updated:

$$\alpha_n = \frac{1}{2} \ln \left( \frac{1 - \varepsilon_n}{\varepsilon_n} \right) \quad (8.5)$$

$$\lambda = \begin{cases} \lambda \frac{1}{2(1 - \varepsilon_n)} & s_n(x) = y \\ \lambda \frac{1}{2\varepsilon_n} & otherwise \end{cases} \quad (8.6)$$

The importance weight  $\lambda$  is passed to the next selector  $s_{n+1}$ . In order to increase the diversity of the classifier pool  $F_n$  and to adapt to changes in the environment the worst weak classifier  $h_{n,m^-}$ , where  $m^- = \operatorname{argmax}_m(\varepsilon_{n,m})$ , is replaced by a classifier randomly chosen from the feature pool  $F$ .



Finally, a strong classifier is computed by a linear combination of  $N$  selectors:

$$H(x) = \text{sign} \left( \sum_{n=1}^N \alpha_n \cdot s_n(x) \right) \quad (8.7)$$

After all online classifiers are constructed, we also use the “Max Wins” voting strategy doing the vehicle classification.

### 8.3 Experiments

In the training phrase, the data set is the image segmentation data, where each class is a vehicle type collected from a  $32 \times 16$  region of a vehicle image. There are four classes to classify: motorcycle, bus, truck, and car. The training set consists of 1000 samples per class while the size of the test set is 400 samples per class. Some training images from all classes are shown in Fig. 8.2. In the test phrase, if all the classifiers obtain detection results, the detection result will be considered to include in the detection rate; if all the classifiers do not obtain detection results or the detection results give the incorrect classification, the detection result will be considered to include in the error rate. We make a comparison of detection rate and error rate with offline SVM multiclass classifiers and online boosting multiclass classifiers using the test dataset. The detection results are shown in Table 8.1. It clearly shows that our method performs better than SVM multiclass learning.



Fig. 8.2 A subset of the training samples for the four classes

Table 8.1 Detection results on test dataset

Classifier	SVM detection rate	SVM error rate	Final detection rate	Final error rate
Motocycle	71	36	92	13
Bus	85	21	98	4
Truck	78	31	96	8
Car	82	27	96	6

More significantly, we create online multiclass classifiers which are suitable for video sequence with small training samples. Some detection results in the video sequences are shown in Fig. 8.3.

We test our system on several video sequences, which consists of more than ten minutes of RGB video taken on city highways during the day. The sequences consist of 150 frames images per class. Table 8.2 shows the results for our detection system, and Fig. 8.4 shows the average detection errors of four classes on the video sequences.

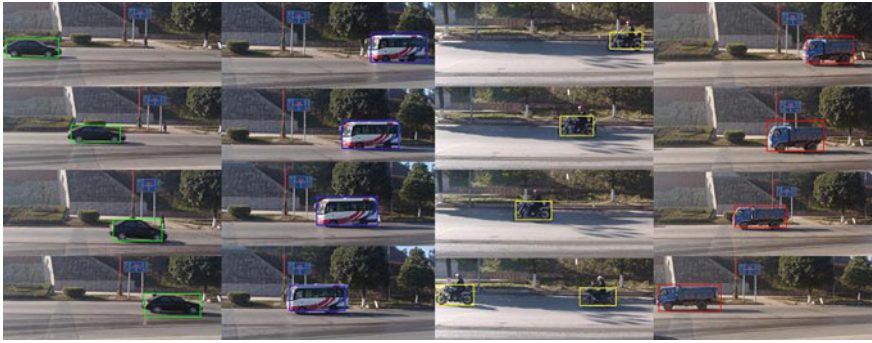


Fig. 8.3 Some detection results in the experimental sequences

Table 8.2 Detection results on video dataset

Classifier	Detected vehicles	Vehicles not detected
Motocycle	14	3
Bus	16	1
Truck	12	0
Car	31	2

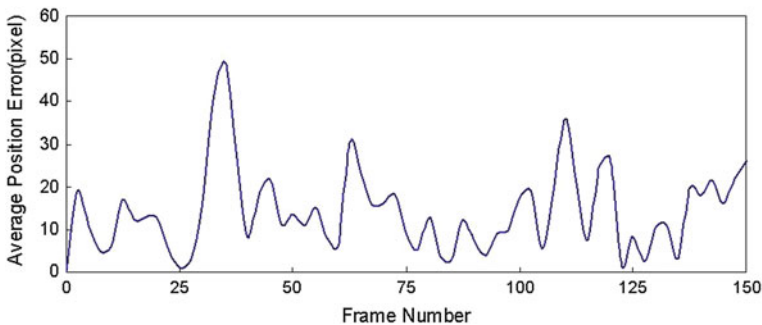


Fig. 8.4 Average detection error on video sequences

## 8.4 Conclusion

We have proposed a real-time vision framework that detects multiclass vehicles in videos. The framework employs a combination of multi-class SVM and online boosting method to analyze the video images. Our framework is able to run in real time with simple, low-cost hardware. Our experimental results demonstrate effective, multiclass vehicle detection in real traffic environments by applying the proposed approach. If new classes of vehicles or unfamiliar environments are encountered, the proposed framework can adapt itself to the changes and detect vehicles successfully.

**Acknowledgments** This work was supported by the Natural Science Foundation of Yunnan Province, China (No. 2011FZ187).

## References

1. Sun ZH, Bebi G, Miller R (2006) On-road vehicle detection: a review. *IEEE Trans Pattern Anal Mach Intell* 28(5):694–711
2. Trivedi MM, Gandhi T, McCall J (2007) Looking-in and looking-out of a vehicle: computer-vision-based enhanced vehicle safety. *IEEE Trans Intell Transp Syst* 8(1):108–120
3. Morris BT, Trivedi MM (2008) Learning, modeling, and classification of vehicle track patterns from live video. *IEEE Trans Intell Transp Syst* 9(3):425–437
4. Chang WC, Cho CW (2010) On-line boosting for vehicle detection. *IEEE Trans Syst Man Cybern Part B (Cybern)* 40(3):892–902
5. Sivaraman S, Trivedi MM (2010) A general active-learning framework for on-road vehicle recognition and tracking. *IEEE Trans Intell Transp Syst* 11(2):267–276
6. Saffari A, Godec M, Pock T et al (2010) Online multi-class LPBoost. In: *IEEE conference on computer vision and pattern recognition*
7. Grabner H, Bischof H (2006) On-line boosting and vision. In: *IEEE conference on computer vision and pattern recognition*
8. Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2(27):1–27
9. Knerr S, Personnaz L, Dreyfus G (1990) Single-layer learning revisited: a stepwise procedure for building and training a neural network. In: Fogelman J (ed) *Neu-roc computing: algorithms, architectures and applications*. Springer, New York

# Chapter 9

## An Object-Level Approach Improved by Quadtree to Dynamic Monitoring of Mining Area Expansion

Liang Huang, Yuanmin Fang, Xiaoqing Zuo and Xueqin Yu

**Abstract** An object-level approach improved by quadtree to dynamic monitoring of mining area expansion is proposed. In order to improve the efficiency and quality of objects acquired from high spatial resolution remote sensing image, multi-scale segmentation combined with quadtree segmentation is used to obtain objects of multitemporal remote sensing images; Then object-oriented image analysis method which takes into account the spatial relationship between ground objects is used in multitemporal remote sensing images to extract mining information respectively; Finally, overlay is use in mining areas extraction respectively, and Inter-erase operation is used to obtain result of mining expansion. Experiments are carried out in remote sensing images from a certain phosphate area of Anning, and the results prove that method was proposed in this paper is feasible and effective.

**Keywords** Mining area expansion · Dynamic monitoring · Quadtree segmentation · Orient-object · Multi-scale segmentation

### 9.1 Introduction

The mineral resources are important natural resources; it is an essential material basis for the development of social production. Mineral resources belong to the non-renewable resources, and reserves are limited. Over the years, the development

---

L. Huang (✉) · Y. Fang (✉) · X. Zuo · X. Yu  
Faculty of Land Resource Engineering, Kunming University of Science and Technology,  
68 Wenchang Road, 12.1 Street, 650093 Kunming, Yunnan, China  
e-mail: kmhuangliang@163.com

Y. Fang  
e-mail: fangyuanmin@126.com

and utilization of mineral resources in China is relatively extensive, some unreasonable exploitations and utilizations, illegal disordered mining and neglect of the mining area ecological environment are serious. These lead to not only the loss of mineral resources, waste and destruction, but pollution of ecological environment around mining and huge life and production losses of the surrounding folks.

There are two main methods to dynamic monitoring the mining development and expansion: (1) field survey; (2) mining expansion dynamic monitoring based on multitemporal remote sensing images. The field survey method can get high accurate monitoring results, but it needs a lot of manpower, especially for complex terrain and large areas of mining. Because remote sensing have many characteristics, such as macroscopic, dynamic, fast and objective, it provides a new technical method for monitoring the change of mine environmental [1]. There are three ways for dynamic monitoring the mining expansion in multitemporal remote sensing images: (1) visual interpretation; (2) pixel-based change detection method; (3) object-level change detection method. Although visual interpretation method can obtain higher monitoring accuracy, but it relies on the expertise and interpret experience of interpretation personnel, and it takes a lot of time and energy, so the timeliness of the information can not be guaranteed. In the past, the low or middle resolution remote sensing images are main image data sources of dynamic monitoring of mining area expansion. Pixel-based change detection method is the main method of change detection based on low or middle resolution remote sensing images; changes are identified by comparing two images that are acquired on the same geographical area at two different times pixel by pixel. Such pixel-based change detection technique is usually divided into three main technological processes: (1) pre-processing; (2) generating a different image by comparing two registered and corrected images pixel by pixel. (3) Analyzing the different image to divide the pixel points into change points or no change points [2]. There are several methods to obtain the difference image, such as image difference method, image ratio method, CVA (Change Vector Analysis) [3–5] and difference method used after transforming the Multitemporal images by PCA (Principle Component Analysis), ICA (Independent Component Analysis) and KICA (Kernel Independent Component Analysis) [6]. Unsupervised change detection on remote sensing images based on Bayesian theory [7] or MRF (Markov Random Field) [8] are the main methods for processing the difference image. With the successful launch of a new generation of high spatial resolution remote sensing satellite, the space structure and surface texture of target ground objects can be more clearly expressed, more sophisticated internal composition of ground objects can be distinguished, the edge information of the ground objects can be seen clearly in the high spatial resolution remote sensing images, so the high spatial resolution remote sensing image becomes the condition and basis of effective mining area expansion dynamic monitoring. It is limited to use pixel-based change detection method to dynamic monitor mining area expansion on high spatial resolution remote sensing image, because there are low separability of each class, huge computation and ambiguity of “change” [9]. In order to solve the problems, the object-level change detection method came into being and widely used. Walter (2004) proposed a

remote sensing image change detection methods based on object-oriented classification. The method classifies objects instead of pixels [10]; Jovanović et al. [11] presented a new method which used object-oriented image analysis techniques for land cover change detection; Bovolo [12] presented a novel multilevel parcel-based approach to detect change in very high resolution multitemporal images; YOU [13] proposed a multi-scale optimization of SAR change detection method, the method used multi-scale segmentation to generate parcels of two images, then cross-entropy was used to calculate the degree of difference between the parcels and obtain the change region. Objects are as units in object-level change detection method, so the intrinsic object size and shape characteristics can be used fully, the separability between the change domain and no change domain can be improved, and so do separability between different ground objects. But acquisition of the objects becomes the difficulty. Now lack of an universal and high precision segmentation algorithm for high spatial resolution remote sensing image is the bottleneck. Comprehensive consideration of the pros and cons of the above methods, An object-level approach improved by quadtree to dynamic monitoring of mining area expansion is proposed. It takes advantage of flexible blocking and high compression ratio of quadtree segmentation [14], it combined with the multi-scale segmentation is used to generate the image objects; Then object-oriented image analysis method which takes into account the spatial relationship between ground objects is used in multitemporal remote sensing images to extract and classify mining information respectively, and the results can be obtained; Finally the mining expansion results can be got by compare the classification results.

## 9.2 Dynamic Monitoring of Mining Area Expansion

### 9.2.1 Proposed Method Scheme

With the continuous improvement of spatial resolution, difficulty in high spatial resolution remote sensing image automatic disposal increases. Mass data and complex details not only make pixel-based change detection inapplicable, but increase difficulty of object-level change detection. Because traditional image segmentation can hardly be used in high spatial resolution remote sensing images directly. In order to solve the problem, an object-level change detection approach improved by quadtree is proposed in this paper. First, preprocessing (radiation correction, geometry correction, image filtering and image registration) is used in two remote sensing images (T1 and T2) acquired on the same geographical area at different times; Second, quadtree segmentation is used in T1 and T2 respectively to obtain original objects, on that basis, multi-scale segmentation is used in T1 and T2 respectively to obtain final image objects; Then, object-oriented image analysis method which takes account the spatial relationship between ground objects is used respectively for image objects classification, meanwhile mining and other

ground object classes can be extracted; finally, results of mining area expansion are obtained through comparing and analyzing results of classification and extraction. The specific process is shown in Fig. 9.1.

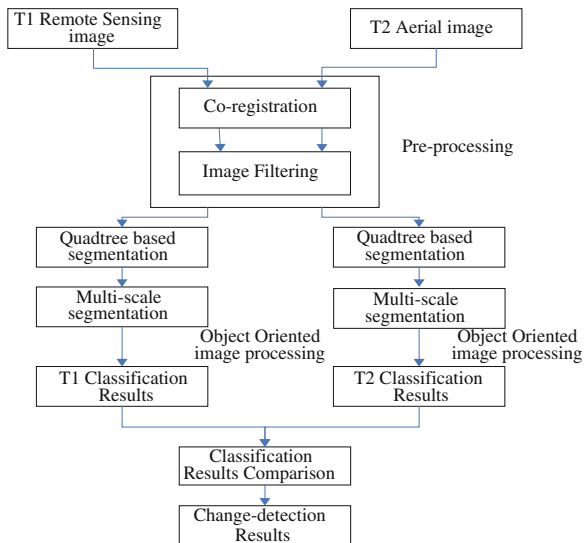
### 9.2.2 Data Source

Two remote sensing images from a certain phosphate area and surrounding areas of Anning are used as data source in experiments. One image (T1) with a spatial resolution of 2.5 m is acquired by the SPOT-5 satellite in 2005, shown as Fig. 9.3a, another (T2) with a spatial resolution of 0.1 m is acquired by photo plane in 2011, shown as Fig. 9.4a. Quality of data is good, and radiation correction and geometric precision correction were carried out in images.

### 9.2.3 Object-Level Change Detection Method

Remote sensing image classification, thematic information extraction and target recognition are main researches of object-level remote sensing image analysis method, but there is very little research on remote sensing image change detection. With continuous improvement of resolution of remote sensing image, object-level change detection has become one of main development trends in the remote sensing image change detection [15]. Image segmentation and change regions extraction are key technologies of object-level change detection method.

**Fig. 9.1** General scheme of the proposed approach



### 9.2.3.1 Image Segmentation

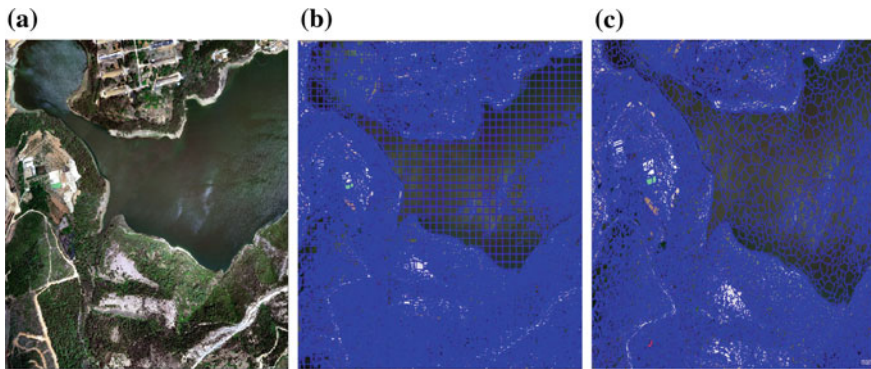
Image segmentation means that specific regions of interest are extracted from remote sensing image, and it is the basis of automatically extracting ground objects with computer, expressing and measuring ground objects. So, Image segmentation becomes one of main research contents of object-level change detection method. At present, multi-scale segmentation algorithm developed in eCognition by Definiens in Germany is the most widely used segmentation algorithm. The algorithm has advantage of that it can reduce average heterogeneity and also increase object homogeneous in more image objects. But it also has some shortcomings, for example, original objects are formed slowly, this can affect overall segmentation efficiency, and inappropriate segmentation scale can cause broken segmentation and over-segmentation. To solve problem of efficiency, quadtree segmentation in remote sensing image is introduced as pre-segmentation in this paper. Quadtree segmentation method is image segmentation algorithm based on uniformity detection. In the algorithm image is expressed as a quadtree, and the original image is expressed as root. In the tree, each tree node has 4 child nodes except leaf nodes, the 4 child nodes mean subblocks in 4 quadrant of original image or image block. The basic idea behind it that an image is subdivided into 4 regions, if a region meets consistency standard, segmentation should be stopped in it; else segmentation is done until each sub-region meets consistency standard. The advantage of this method is high flexibility and fast computation speed [16].

In the paper, quadtree segmentation was first used in image to obtain original objects, then multi-scale segmentation was used after the first segmentation, and finally image objects were obtained. To compare efficiencies and qualities of multi-scale segmentation algorithm combined with quadtree segmentation and traditional multi-scale segmentation algorithm, experiments were carried out in local image (Fig. 9.2a). In experiments, segmentation scale was chosen as 100. The run time of multi-scale segmentation algorithm combined with quadtree segmentation was 1067.124 s, run times of quadtree segmentation and multi-scale segmentation were 42.282 s and 1024.842 s; but run time of traditional multi-scale segmentation algorithm was 1665.208 s. Results of multi-scale segmentation algorithm combined with quadtree segmentation and multi-scale segmentation algorithm were shown respectively in Fig. 9.2b, c. The results of experiments proved that efficiency of multi-scale segmentation algorithm combined with quadtree segmentation was improved and segmentation quality could meet requirement of subsequent change regions detection.

### 9.2.3.2 Change Region Extraction

There are two change region extraction methods: one is post-classification comparison; another is extracting after direct comparison between objects. Because of different data source and huge data, direct computing can not be implemented effectively between region objects using current algorithms. Post-classification

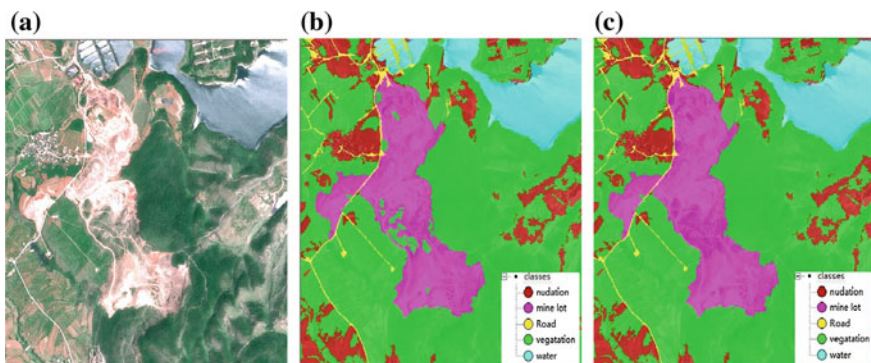




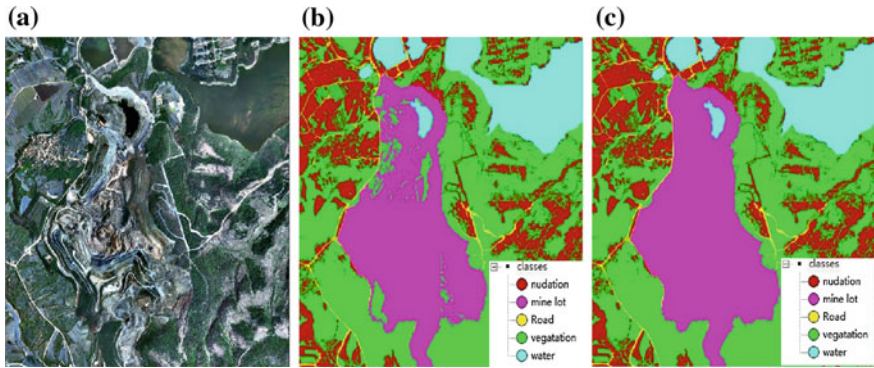
**Fig. 9.2** Image segmentation: **a** original image; **b** results of multi-scale segmentation algorithm combined with quadtree segmentation; **c** results of multi-scale segmentation

comparison depends on high-precision classification results, however high-precision classification results are often difficult to obtain. So, now efficiency of post-classification comparison is still much higher than efficiency of direct comparison between mass objects. To obtain high-precision classification results, object-oriented image analysis method which takes into account the spatial relationship between ground objects is used in T1 and T2 for images classification. The results are shown in Figs. 9.3 and 9.4, Figs. 9.3a and 9.4a are the results including mining, vegetation and water. The following classification strategies were carried out in two different temporal images:

- (1) Vegetation and water: First of all, NDVI was used to differentiate vegetation areas and nonvegetated areas, in this step water was recognized as vegetation; and then, Mean Blue was used to separate water from vegetation areas through spectrum difference between vegetation and water. For example, NDVI was chosen as  $(-0.9, 0)$  to differentiate vegetation areas (include water) and nonvegetated areas, Mean Blue  $> 100$  was set to separate water from vegetation areas.



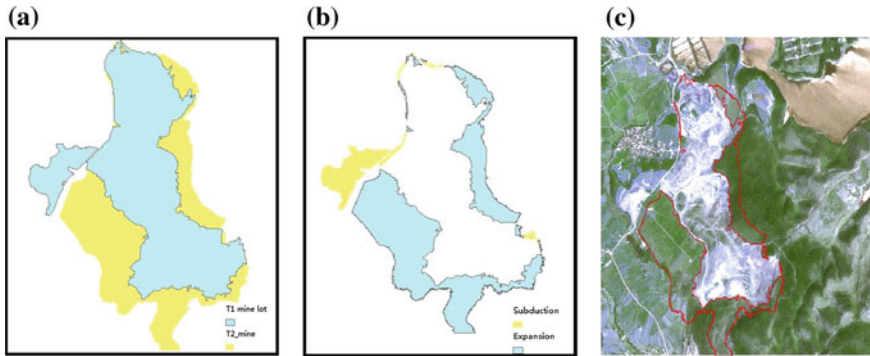
**Fig. 9.3** Classification results of T1 image: **a** T1 image; **b** mine include vegetation; **c** mine



**Fig. 9.4** Classification results of T2 image: **a** T2 image; **b** mine include vegetation and water; **c** mine

- (2) Road network: Part of roads was difficult to extract in remote sensing image T1 for that resolutions of remote sensing images T1 and T2 are different, so only the main roads would be extracted in images (T1 and T2). In order to best use the road long/width ratio, and taking into account that the spectral characteristics of the roads, bare ground and mining are similar, firstly the roads segmented were merged artificially, and then road network was extracted by road long/width ratio.
- (3) Bare land and mining area: Because the spectrum, shape and texture of bare land and some mining areas were extremely similar, mining areas were difficult to extract only by spectrum, texture and geometric feature. In the paper, the spatial relationship between ground objects was introduced to mining extraction. Mean Red, Mean Green and Mean blue of greater than 230 were set to extract some mining areas for that Mean Red, Mean Green and Mean Blue of phosphorite were greater than 230; relationship with adjacent objects of mining was used to separate from areas, in which spectrum of mining and bare land were similar. For example, according to distance relations in “Relations to Neighbor Objects”, Distance to Mine lot of greater than 30 pixels was set to extract mining in T1 through classification. Semi-automatic or human-interactive method was used to extract through classification in mining areas, whose spectrum, shape and texture were similar to them of bare land. The relationship of coincidence boundary in all directions of vegetation or water and mining was used to extract vegetation and mineral bath from mining areas. For example, mineral bath and vegetation in mining areas were extracted as mining by border relationship in “Relations to Neighbor Objects” through setting Border to Mine lot of greater than 0 pixel and Border to Vegetation/Water of less than 0 pixel.

After the classification results of T1 and T2 were obtained, overlay (Fig. 9.5a) was implemented in the output mining areas of T1 and T2 using Arcgis9.3. Then results (Fig. 9.5b) of mining area expansion were acquired through inter-erase.



**Fig. 9.5** Change analysis: **a** mining area overlay of T1 image and T2 image; **b** mining area expansion and subduction; **c** expansion area overlay of T1 image

### 9.2.3.3 Results Analysis

Mining areas are acquired through analyzing the results of mining information extraction in T1 and T2 respectively. There are two mining areas in research area of 2005, and areas of them are  $69078.125 \text{ m}^2$  and  $673218.750 \text{ m}^2$ . Mining of 2011 has an area of  $1114850.481 \text{ m}^2$ . Expansion area of  $455341.964 \text{ m}^2$  is acquired from 2005 to 2011 through inter-erase, two mining areas of 2005 reduce  $69078.125 \text{ m}^2$  and  $13710.230 \text{ m}^2$  respectively (Fig. 9.5b). As Fig. 9.5c shows, expansion area is main forest in 2005, and mining area expansion destroys local forest. However, mining reduction area is forest, which was mining before; Figs. 9.3b and 9.4b show that exploited mining are gradually transformed into vegetation.

## 9.3 Conclusion

Greatly improve the efficiency of the high spatial resolution remote sensing image segmentation by quadtree segmentation combined with multi-scale segmentation methods. The object-oriented image analysis method which takes into account the spatial relationship between ground objects can take full advantage of various kind of spatial relationship has greatly improved the accuracy of the mining classification and extraction. In the basis of the two, mining information were separately extracted, then the mine information of multitemporal remote sensing images were processed by inter-erase, and finally the mining expansion results were obtained. The expansion results prove that method was proposed in this paper is feasible and effective. Because of different objects with same spectrum, the capability of applying remote sensing images on mining classification is reduced. Due to interference of different objects with same spectrum, part of the misclassification

occurred. But correct the misclassification is difficult to achieve a fully automated, partially corrected using a semi-automatic way, and thus how to better automatic extract the mining information need to be further study.

**Acknowledgments** This research work was supported by the National Science Foundation of China (NO. 41061043) and Department of Education Research Fund of Yunnan Province (No. 2011J075).

## References

1. Qi XY, Yan MX (2007) Dynamic monitoring of mining area expansion based on multitemporal remote sensing images. *Remote Sens Land Resour* 3:85–88 (in Chinese)
2. Bruzzone L, Prieto DF (2002) An adaptive semi parametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images. *IEEE Trans Image Process* 11(4):452–466
3. Wen XP, Yang XF (2009) Change detection from remote sensing imageries using spectral change vector analysis. In: *Asia-Pacific conference on information processing*, pp 189–192
4. Mura MD, Benediktsson JA, Bovolo F et al (2008) An unsupervised technique based on morphological filters for change detection in very high resolution. *IEEE Geosci Remote Sens Lett* 5(3):433–437
5. Bovolo F, Bruzzone L, Marchesi S (2007) A multiscale technique for reducing registration noise in change detection on multitemporal VHR images. In: *International workshop on the analysis of multi-temporal remote sensing images*, pp 1–6
6. Marchesi S, Bruzzone L (2009) ICA and kernel ICA for change detection in multispectral remote sensing images. In: *IEEE international geoscience and remote sensing symposium*, pp 980–983
7. Chen X, Dai Q, Ma JW, Li XW (2005) Application of bayesian network classification to remote sensing change detection. *J Beijing Normal Univ (Nat Sci)* 41(1):97–100 (in Chinese)
8. Gabriele M, Elena A, Sebastiano BS (2009) A contextual multiscale unsupervised method for change detection with multitemporal remote-sensing images. In: *2009 ninth international conference on intelligent systems design and application*, pp 572–577
9. Huo CL, Cheng J, Lu HQ, Zhou ZX (2008) Object-level change detection based on multiscale fusion. *Acta Autom Sin* 34(3):251–257 (in Chinese)
10. Walter V (2004) Object-based classification of remote sensing data for change detection. *ISPRS J Photogram Remote Sens* 58:225–238
11. Jovanović D, Govedarica M, Đorđević I, Pajić V (2010) Object based image analysis in forestry change detection. In: *2010 IEEE 8th international symposium on intelligent systems and informatics*, pp 231–236
12. Bovolo F (2009) A multilevel parcel-based approach to change detection in very high resolution multitemporal images. *IEEE Geosci Remote Sens Lett* 6(1):33–37
13. You HJ (2011) SAR change detection by multi-scale segmentation and optimization. *Geom Inf Sci Wuhan Univ* 36(5):531–534 (in Chinese)
14. Gao W, Liu XG, Peng P, Chen QH (2010) An improved method of high-resolution remote sense image segmentation. *Earth Sciences—J China Univ Geosci* 35(1):421–425 (in Chinese)
15. Zhang HS (2010) The research of object-based remote sensing change detection for coastal surface. *Zhejiang University, Hangzhou*, pp 23–24 (in Chinese)
16. Zhou SL, Liang D, Wang H, Kong J (2010) Remote sensing image segmentation approach based on quarter-tree and graph. *Comput Eng* 36(8):224–226 (in Chinese)

# Chapter 10

## Heuristic Optimization Algorithms for Solving MRMPT

Chunhua Meng, Hongguo Wang, Zengzhen Shao and Yanhui Ding

**Abstract** Multi-vehicle ride matching problem with transfers (MRMPT) studies how to match passengers as much as possible to vehicles at the basis of MRMP which has no transfer mechanism. The passengers in this problem are classified according to their needs. The first-level passengers have been matched in the basic MRMP. And what we are solving is matching the second-level passengers in MRMPT. To improve the riding rate, a heuristic algorithm based on ant optimization for the MRMPT is proposed. The algorithm is divided into three steps: Finding starting and ending sets; optimizing the routes by ant optimization; Trimming the vehicle routes. Simulation experiment shows that the algorithm can achieve riding rate of 80 %. The results show clearly that the algorithm can find the matching routes with high efficiency and low cost.

**Keywords** MRMPT · Heuristic algorithm · Transfers · Passengers classification

### 10.1 Introduction

Traffic congestion has always been a big problem for many countries. In order to deal with the problem of traffic congestion, the “free-rider” phenomenon came into being. There are some carpool information publishing platforms for travellers selecting vehicles to ride on the internet. But these are mere simple selecting artificially. Thus it is vital to research on the algorithm about vehicle ride matching problem.

Transfers and passengers classification have seldom been addressed in the literature on multi-vehicle ride matching problem. References [1–3] studied the PDP

---

C. Meng (✉) · H. Wang · Z. Shao · Y. Ding  
School of Management and Engineering, Shandong Normal University,  
No. 88 East Wenhua Road, Jinan, China  
e-mail: meng1662@163.com

problem with transfers. Sophie N [4] proposed a solution of dial-a-ride problem based on the heterogeneous users. Hame [5] solved single-vehicle dial-a-ride problem using an adaptive insertion algorithm. There are also many literatures that solve routing problems using ant algorithm [6, 7], which showed that the ant algorithm has a better ability to find excellent routings. Shao et al. [8] proposed a two-stage clustering algorithm to solve the static MRMP with time windows. Based on this framework, we studied MRMPT increasing transfers and passengers classification mechanism.

## 10.2 MRMPT Model with Passengers' Classification

### 10.2.1 Description of the Problem

Within a region, a car fleet  $F$  consisting of  $m$  vehicles, can provide riding passengers on the basis of their own travel needs. And there are  $n$  passengers. The brand, spare capacity and travel speed of the vehicles are different. Vehicles information and needs information are certain and constant. The up and off stations of the vehicles and passengers are certain and have corresponding fixed time windows. The object is matching the  $n$  passengers as much as possible to the  $m$  vehicles. This is the basic MRMP problem.

In this work, MRMPT with passengers' classification considers different priority of passengers. Passengers are divided into two levels. The first level of passengers belongs to the basic multi-vehicle matching problem areas. The second hierarchical passengers can transfer and have no strict time window constraints. What we are to solve is finding routes for the second hierarchical passengers based on the vehicles' routes with the first hierarchical passengers. The transfer points which are in the set of intersections of the previous vehicle routes are not fixed. Thus, it won't produce inflexibility because of the increase of fixed transfer points.

### 10.2.2 Problem Formalization and Symbol Definitions

Assume that the number of the intersections of the vehicle routes  $k$ . The number of passengers of the first priority is  $n-d$ . Thus, the remaining quantity having second priority passengers is  $d$ . Let  $F$  be vehicle set:  $F = \{1, 2, \dots, m\}$ ;  $P$  be passengers set:  $P = \{1, 2, \dots, n\}$ ;  $D$  be the second level passengers set:  $D = \{1, 2, \dots, d\}$ .

Assuming  $F^+$  as set of vehicles' up points,  $F^-$  as set of vehicles' off points,  $P^+$  as set of passengers' up points,  $P^-$  as set of passengers' off points,  $V$  as the set of all points on the path.  $Tr$  as set of all routing crosspoints:  $Tr = \{1, 2, \dots, k\}$ ;

Defining the remaining capacity of the vehicle  $k(k \in F)$  as  $Q_k$ , The Traveling time that vehicle  $k$  required from point  $i$  to point  $j$  as  $t_{ij}^k$ .

$F_T$  is the number of transfer points;  $T_d$  is time loss resulting from transferring.

### 10.2.2.1 Objective Function

The objective is to improving the riding rate, and then reducing total costs (including costs from transferring).

$$\max \sum_{j \in F} \sum_{x \in S} \sum_{y \in V} X_{x,y}^j \quad (10.1)$$

$$\min \sum_{j \in F} (fc_j + \sum_{u=1}^{U_j} ac_j \cdot s_u^j \cdot q_{x_{u-1}}^j) + \sum_{m \in D} F_T \cdot T_d \quad (10.2)$$

A binary variable  $X_{x,y}^j$  is assigned to each arc  $\langle i, j \rangle$  and vehicle  $j$ , so that  $X_{x,y}^j$  is 1 if vehicle  $j$  travels through arc  $(i, j)$ , and otherwise is 0.  $x, y \in V$ ,  $x \neq y, j \in F$ .

$fc_j$  is the fixed costs of vehicle  $j$ . The variable cost is  $ac_j \cdot s_u^j \cdot q_{x_{u-1}}^j$ .  $ac_j$  is the costs that vehicle  $j$  increases one unit goods in every unit mileage.  $s_u^j$  is the distance between station  $x_u$  and  $x_{u-1}$ .  $q_{x_{u-1}}^j$  is the goods capacity when arriving at station  $x_u$

### 10.2.2.2 Constraints

(a) Time windows constraints

$$T_x^j \leq l_x, \quad x \in P, j \in F \quad (10.3)$$

$$T_x^j - T_x^i \leq T_d, \quad x \in D, i, j \in F \quad (10.4)$$

(b) Vehicle capacity constraint

$$q_x^j = \text{init}q_j, \quad x \in F^+, j \in F \quad (10.5)$$

$$q_x^j = \text{init}q_j, \quad x \in F^-, j \in F \quad (10.6)$$

$$q_x^j \leq Q_j, \quad x \in P, j \in F \quad (10.7)$$

(c) Visit order constraint

$$T_x^j + RT_x \leq T_{m+n+x}^j, \quad x \in N^+, j, j' \in F \tag{10.8}$$

(d) Trasfer times constraint

$$F_T \leq r \tag{10.9}$$

### 10.3 Heuristic Optimization Algorithm Based on Ant Colony for MRMPT

The ant colony algorithm is a randomized algorithm for solving complex combinatorial optimization problems. Bullnheimer [9] showed that the ants algorithm is competitive when compared with other metaheuristics such as Tabu Search, Simulated Annealing and Neural networks. The most important two parts in the ants—pheromone updating stragety and selecting stragety, will be discussed in detail later in the paper.

The flow of the algorithm based on ant colony is shown in Fig. 10.1

This three-step in looking for every customer point of transfer route is serial, in order to prevent confusion between routes.

#### 10.3.1 Step 1: Finding Out the Starting and Ending Sets

Let  $s$  be start point of the new passenger,  $e$  be end point. Assuming starting set as  $S: \{S_{i1}, S_{i2}, \dots, S_{ik}\}$ , ending set as  $E: \{e_{i1}, e_{i2}, \dots, e_{ik}\}$ .  $i \in P, k \in F$ . When finding S and E, we used a distance clustering. Seting a proper parameter  $d$ . Let's see how to find S as an example. With start point as the center, finding all points that in the distance of  $d$  with  $s$  in the original routes net. These points can be added to S. In the same way, we can find the set E.

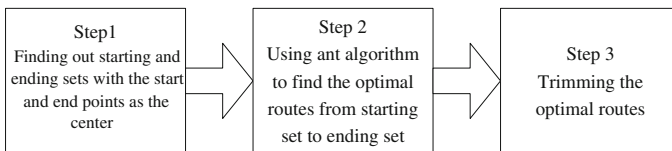


Fig. 10.1 Algorithm processes



### ***10.3.2 Step 2: The Ant Colony Algorithm to Find the Optimal Transfer Route***

S can be regarded as the anthill in the ants colony, and E is regarded as the food source. The original routes of the vehicles are the routes net of the new problem. Ants forage on the routes net from S to E. And in the end, we can find an optimal routes of all ants. In the process, These factors must be considered: time window of the vehicles and their passengers on the vehicle, the capacity constraint of the vehicles.

#### **10.3.2.1 One-Way Ants**

The algorithm is shown in the following steps;

---

One-way ant colony algorithm steps

**Initialization**

Structuring pheromones matrix, and the same number  $m$  ants are placed on the every point in the starting set  $S$ ;

**Repeat**

**Repeat**

From a starting point in the  $S$ , ants move to the next point according to the transition rules, and the point was added to the corresponding ant's taboo table;

If transfer times reach  $r$ , the ant will die;

If the ant reaches any point of the ending set E, recording the route and update the optimal solution. Then enter the next starting point for optimization;

**Until** all ants complete their foraging;

According to pheromone amending rules, update pheromone of the path;

**Until** meet the constraint of the max iterations  $N_c$ ;

---

The optimum of the ants is the least time instead of shortest distance. This is because that in the routes net composed of vehicles the speed is not constant. This paper presents two kinds of ant colony optimization algorithm and compares them with each other. One kind is one-way ants in which all ants start at starting set and end at ending set. The other kind is two-way ants in which part of ants start at starting set and end at ending set and other parts start at ending set and end at starting set.

---

Two-ant ant colony algorithm steps

**Initialization**

Structuring pheromones matrix, and an equal number  $m/2$  ants were placed on the starting set  $S$  and the ending set  $E$  point;

**Repeat**

**Repeat**

From a starting point in the  $S$  or in the  $E$ , ants move to the next point according to the transition rules, and the point was added to the corresponding ant's taboo table;

If transfer times reach  $r$ , the ant will die;

If any ants from  $S$  encounter the ants from  $E$  or the ant reach its food source, recording the routes and update the optimal solution. And According to pheromone amending rules, updating pheromone of the path

**Until** all ants complete their for aging, or there are no path to go;

**Until** meet a number of iterations  $Nc$  bound;

---

### 10.3.2.2 Two-Ant Colony

The steps are as follows:

#### 10.3.2.3 Pheromone Updating and Transition Probabilities

In the ant colony algorithm, the pheromone updating and the transition probability of ant's path selection are essential. When selecting routes ants refer to the time instead of distance. So the pheromone update strategy as well as transition probabilities are improved. The improvement is given by the following method:

(a) Pheromone updating

Two factors should be taken into account when selecting route: transfer times and travel time. The updating rule is given according to the two factors.

The increasing pheromone is as follows:

$$\Delta R = R\lambda \left( \frac{n-f}{n} \right)^{\alpha 1} \left( 1 + \frac{1}{\sqrt{t^{\beta 1}}} \right)$$

- $\lambda$  is the importance of selecting vehicles, mainly determined comprehensively according to the remaining capacity as well as the time window of the vehicle. The general value of 0.8–1.2;
- $n$  is the maximum number of transfers
- $f$  is the transfer times that the current trip need;
- $\alpha 1$  is the coefficient that transfer times influence on path selection;
- $t$  is travel time that the ant required;
- $\beta 1$  is the coefficient of travel time influence on the path selection.

When ants complete foraging, pheromone strength on the path that ants don't pass will gradually weakened as time goes by. The dissipation rate of the pheromone is  $\rho = 10\text{--}30\%$ . After some time, there will be a certain degree of pheromone dissipation on all the routes.

Pheromone update formula is as follows:

$$R_{new} = \begin{cases} (1 - \rho) R_{old} + \Delta R; \\ (1 - \rho) R_{old}; \end{cases}$$

### (b) Transition probabilities

In the foraging process, ants can't select the path they have travelled. So the taboo table  $T_k$  that ants can't travel for every ant can be established.

$$P_{ij}(k) = \begin{cases} \frac{R_{ij}^\alpha \times \eta_{ij}^\beta}{\sum_{h \notin T_k} R_{ih}^\alpha \times \eta_{ih}^\beta}; & j \notin T_k \\ 0; & \text{Others} \end{cases}$$

- $P_{ij}(k)$  is the probability that the  $k$ th ant chooses the edge  $(i, j)$
- $R_{ij}$  is the strength of the edge  $(i, j)$  of the pheromone
- $\eta_{ij}$  is the visibility of the edge  $(i, j)$  ·  $\eta_{ij} = 1/t_{ij}$

### 10.3.3 Step 3: Trimming the Optimal Routes

Trimming the optimal routes that ant colony algorithm found is matching the up and down station to the original routes of the vehicles. For example, let  $s$  and  $e$  be the passenger's up and down station. Ant colony algorithm's optimal route is:  $\langle r_1-r_2-r_3-r_4-r_5-r_6-r_7 \text{ (transfer)}-r_8-r_9 \rangle$ . That is  $r_1$  and  $r_9$  aren't necessarily  $s$  and  $e$ . If route  $\langle r_1-r_2-r_3-r_4-r_5-r_6-r_7 \text{ (transfer)} \rangle$  belongs to vehicle  $i$ , and route  $\langle r_7 \text{ (transfer)} -r_8-r_9 \rangle$  belongs to vehicle  $j$ , the passenger needs to ride vehicle  $i$  first and then transfers to vehicle  $j$  to reach the destination. Assuming that the start and end station of the passenger are fixed, the vehicles have to trim their routes to carry the passenger on condition that they don't change the original passengers' up and down station.

When trimming, we do insertion sort for the vehicle's route. The needs of first-level passengers have to be noticed. The trimmed routes must meet first-level passengers' up and down stations and their time windows.

First, Find out the vehicle that include the subroutes before the first transfer point. Then add the up station  $s$  of the new passenger to the vehicle's route. The method is: Find out the passengers' points in the vehicle's route before the first

transfer point. Put these passengers' points and transfer points in a list. The insert the up station  $s$  into the list. Then sort the list to find the order of the minimum cost. After the sorting, insert the intermediate node into the order to get the new route of the vehicle. For example, the original route of the  $i$ th vehicle is  $\langle i_1-i_2-\underline{r_1-r_2-r_3-r_4-r_5-r_6-r_7-i_3-i_4} \rangle$ . The trimming route is  $\langle i_1-i_2-\underline{s-r_1-r_2-r_3-r_4-r_5-r_6-r_7-i_3-i_4} \rangle$  (the underline part is the subpath of the new passenger).

For the last transfer point, using the same method, insert the off station  $e$  to the route of the vehicle that include the subroute behind the last transfer point. For example, The original route of the  $j$ th vehicle is :  $\langle j_1-\underline{r_7-r_8-r_9-j_2-j_3-j_4-j_5-j_6} \rangle$ , the trimming route is  $\langle j_1-\underline{r_7-r_8-r_9-j_2-e-j_2-j_3-j_4-j_5-j_6} \rangle$ .

For the middle transfer points, only consider if the transfer cost influence the passengers' time windows.

In this way, a final route of the new passenger is acquired:  $\langle s-r_1-r_2-r_3-r_4-r_5-r_6-r_7$  (transfer)  $-r_8-r_9-j_2-e \rangle$ .

## 10.4 Experimental Results Show

The experimental data is based on the experimental data and results of literature [8]. Assuming that there are  $m = 10$  vehicles that can provide riding service, and there are 30 first-level passengers. In the first-stage matching, 28 passengers have been ridden. Based on the data, adding a second-level passenger at a time. Five s-level passengers are added in series. The contents of the experiment include: finding the optimal transfer routes and the comparison between the two kinds of ant algorithm.

### 10.4.1 Parameter Settings

The experiment parameter settings are shown in Table 10.1.

The distance parameter  $d$  is the first quartile of the distance between the points in the database, which is actual distance instead of straight-line distance.

**Table 10.1** Parameters set

Parameter name	Parameter values	Parameter name	Parameter values
$d$	150	$\alpha$	1
$m$	15	$\beta$	1
$\rho$	0.15	$\lambda$	(0.8-1.2)
$\alpha$	0.8	f	1
$\beta$	0.5	Nc	20
TL	30		

**Table 10.2** Performance comparison between one-way ant colony and two-way ant colony

Passenger num	One-way /Two-way optimal solution	Oneway / two-way transfer times	One-way /two-way running time (in s)
31	17, 16, 15, 33, 23, 35, 36, 42, 56/ 17, 16, 15, 33, 23, 35, 36, 42, 56	1/1	528/ 495
32	8, 13, 35, 36, 42/ 8, 13, 35, 36, 42	2/2	1860/ 1320
33	26, 31, 51/ 26, 31, 51	0	2690/ 2360
34	12, 36, 42, 56, 57, 59, 60, 61, 64, 80/ 12, 36, 42, 56, 57, 59, 60, 61, 64, 80	1/1	2960/ 2255
35	26, 31, 51, 91, 90, 79/ 26, 31, 51, 91, 90, 79	1/1	2771/ 1990

### 10.4.2 Analysis of Experimental Results

Experimental results are shown in Tables 10.2 and 10.3.

As can be seen from Table 10.2, the results of one-way and two-way ant colony optimization are the same, which is largely due to the reason that the number of ants in the ant colony parameters and the number of iterations are set sufficiently large enough to find the optimal transfer path for ants colony every time. However, one-way ant colony which used less running time is more efficient than two-way.

Table 10.3 shows that the 32th passenger has no solution, which is because the transfer times are limited. And even if the transfer times are not limited, it doesn't meet the time window constraint of the 8th vehicle.

Figure 10.2 describes the transfer route of passenger 31, whose up station is 20, down station is 58. The optimum transfer route by ant algorithm is < 17, 16, 15, 33, 23, 35, 36, 42, 56 >. And from the ant algorithm, we can know the passenger need ride vehicle 9 first and then ride transfer at station 36 to ride vehicle 8. Because the up and down station of passengers are fixed, we need trim the routes of vehicle 8 and vehicle 9 to carry the passenger 31.

The final route of the passenger 31 is < 20, 17, 16, 15, 33, 23, 35, 36, 42, 56, 57, 58 >. For vehicle 9, the route before carrying passenger 31 is < 5, 18, 17, 16, 15, 33, 23, 35, 36, 37, 38, 39 >, and after trimming, the route is < 5, 18, 19, 20,

**Table 10.3** Second level customer path fine-tuning after solution

Passenger num	Passengers' solution routes	Transfer points	Vehicles ridden	Total cost
31	20, 16, 33, 36, 42, 58	36	9-8	5430
32	No solution			
33	6, 17, 24, 26, 31, 74, 75, 102	No	4	5534
34	13, 36, 42, 41, 58, 59, 61, 64, 80, 82	64	8-3	5892
35	27, 24, 26, 31, 51, 91, 90, 79	51	4-10	5690

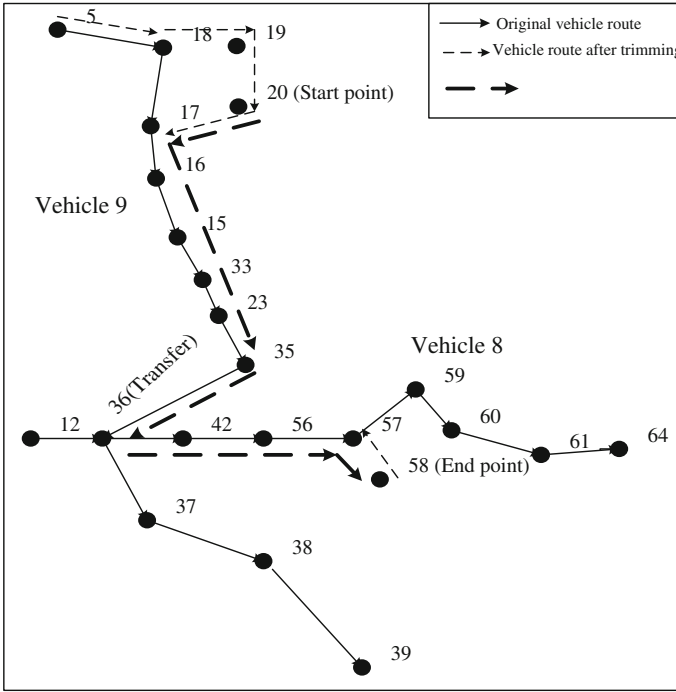


Fig. 10.2 Transfer path of passenger 31

17, 16, 15, 33, 23, 35, 36, 37, 38, 39 >. For vehicle 8, the route before carrying passenger 31 is < 12, 36, 42, 56, 57, 59, 60, 61, 64 >, and after trimming, the route is < 12, 36, 42, 56, 57, 59, 60, 61, 64 >.

### 10.5 Conclusion

This paper researched on the MRMP with transfers and passengers' classification. And the three-step solution based on ants algorithm is proposed: finding starting and ending set; ant algorithm optimization; trimming the routes. Finally, the experiment showed that the algorithm can solve the second matching problem, and has better effect. However, some passengers can't be matched, which is due to the time window and capacity constraints of the vehicles and the first-level passengers. For the future research, two aspects can be considered: (1) The MRMP model this paper proposed is still in static stage. In future work, we consider adding the influence of the traffic flow to realize dynamic MRMP. (2) The parameters combinations in the algorithm are to be further improved. A large number of experiments is required to optimize the algorithm.

## References

1. Cortés CE (2010) The pickup and delivery problem with transfers Formulation and a branch-and-cut solution method. *Eur J Oper Res* 200(3, 1):711–724
2. Jia YJ, Gu HY, Xi YG (2004) Quick taboo search algorithm for solving PDPTW problem. *Control Decis* 19(1):57–60 (in Chinese)
3. Li L, Gu HY, Chen J (2003) Insertion heuristic algorithm in complex PDPTW problem. *Comput Eng* 16(2):65–69 (in Chinese)
4. Parragh SN (2011) Introducing heterogeneous users and vehicles into models and algorithms for the dial-a-ride problem. *Transp Res Part C Emerg Technol* 8, 19(5):912–930
5. Häme L (2011) An adaptive insertion algorithm for the single-vehicle dial-a-ride problem with narrow time windows. *Eur J Oper Res* 2, 209(1): 11–22 (2010.08.021)
6. Li WY, Wang W, Chen XW (2004) Bus travel transit path based on ant algorithm. *J Traffic Tansp Eng* 4(4):102–105 (in Chinese)
7. Balseiro SR (2011) An ant colony algorithm hybridized with insertion heuristics for the time dependent vehicle routing problem with time windows. *Comput Oper Res* 38(6):954–965
8. Shao ZZ, Wang HG, Liu H et al Heuristic optimization algorithms of multi-carpooling problem based on two-stage clustering. *J Comput Res Dev*. To be published (in Chinese)
9. Bullnheimer B, Hartl RF, Strauss C (1999) An improved ant system algorithm for the vehicle routing problem. *Ann Oper Res* 89:319–328

# Chapter 11

## A Delay-Based Analysis of Multiple Bottleneck Links of End-to-End Paths in the Internet

Jingang Liu, Wei Peng, Yonglei Yang and Zhijian Huang

**Abstract** Measurement and analysis of bottleneck links play an important role in improving the network quality of service (QoS) and preventing network attacks in the Internet. Existing methods usually treat the link with the smallest available bandwidth or the largest delay as the bottleneck link, without considering multiple bottleneck links in an end-to-end path. In this paper, we propose a new approach to measure and analyze bottleneck links based on path delay. We design a parallel active measurement framework to measure the path delays of many destinations simultaneously. Then an algorithm to identify multiple bottleneck links is proposed using the Ward data clustering method. Experiments are conducted to test the algorithm by measuring 10 different destinations in the Internet for 14 days. Using the proposed approach, we have found that bottleneck links are mainly few constant links which are in the intermediate of end-to-end paths or near the destinations. Furthermore, the results have shown that the number of intra-domain bottleneck links takes a large portion in most cases, which hints that the performance of end-to-end paths may be greatly influenced by iBGP routing. Besides, the results have also demonstrated that the intercontinental links in an anonymous system (AS) incline to be bottleneck links in end-to-end paths.

**Keywords** Bottleneck link · Delay · Network measurement · Data clustering

### 11.1 Introduction

With the rapid development of the Internet, a variety of network applications have emerged and people's life is heavily depended on the Internet. Network performance has become one important problem for the Internet Service Providers (ISPs)

---

J. Liu (✉) · W. Peng · Y. Yang · Z. Huang  
School of Computer, National University of Defense Technology Changsha,  
Hunan 410073, China  
e-mail: taobao.qin@gmail.com



and researchers. Bottleneck links represent the handicaps that prevent the improvement of network performance. It has become an important research topic on how to detect and eliminate them.

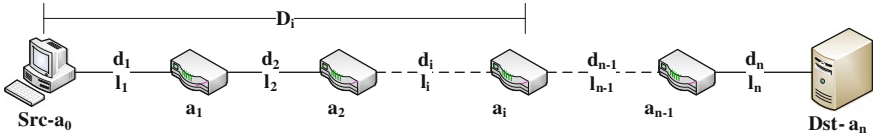
Currently, the most widely-used methods use available bandwidth in the measurement and analysis of bottleneck links [1, 2]. The available bandwidths of links in an end-to-end path are estimated using active probing method and the link with the smallest available bandwidth is identified as the bottleneck link normally. Typical software tools include Bfind [3] and Pathneck [4]. In [5], a detection algorithm is proposed based on statistical methods to detect the presence of bottleneck links by examining strong frequencies in aggregate data traffic. It needs multiple monitoring points and assumes that the traffic through the bottleneck link is dominated by packets with a common size, thus its applicability is limited. In [6], a statistical model for packet transmission probability is proposed to detect network bottleneck links. The method is based on the assumption of having knowledge about network structure. There are few approaches on measurement and analysis of bottleneck links using network delays.

Some work have been done on measuring and analysis of the characteristics of network delays [7–9]. In this paper, we focus on the study of the bottleneck links determined by network delays. By this study, we can find out bottleneck links more accurately and the impact of network delays. A link in an end-to-end path becomes a bottleneck link if it satisfies the following two conditions:

- (1) The link delay should be larger than a threshold value. In other words, the link delay should be large enough to influence the network performance;
- (2) The link delay is obviously greater than most of other links in the path.

There may be more than one bottleneck link in an end-to-end path. In previous research work, the link with the maximal delay is usually considered as the bottleneck link and only one bottleneck link is checked. In [10], the authors analyze the characteristics of link delays along some end-to-end paths located among some servers which are deployed in the Internet for measurement. In [11], the raw giant data samples authorized by CAIDA [12] are studied. The former identifies bottleneck links by calculating the largest link delay on the path. It does not solve the problem of multiple delay bottlenecks in an end-to-end path. The later identifies delay bottlenecks using predefined delay ranges. It has the ability to identify multiple delay bottlenecks in an end-to-end path. However, there lacks a delay range which is commonly accepted as criteria of bottleneck links.

To identify multiple bottleneck links in an end-to-end path simultaneously, we propose a new approach using data clustering methods. Specifically, we use the WARD clustering method [13, 14] to analyze the link delay data and the links within the group with the largest delays are identified as bottleneck links. To gather data to test the method, we measure link delays in the Internet using an improved traceroute program which probes many destinations parallelly. The experimental results have shown that the proposed method can accurately locate multiple delay bottlenecks in most cases. Besides, we find that a large portion of bottleneck links are distributed in the middle of paths or near the destinations. And



**Fig. 11.1** Illustration of an end-to-end path

bottleneck links may come from intra-domain links which locate in the interior of some autonomous systems (ASes), which suggests that the routing performance of some transit ASes has great impact on the quality of service (QoS) of the Internet.

The rest of the paper is organized as follows. We describe the measurement method and the method of identifying and analysis of multiple bottleneck links in [Sect. 11.2](#). Experiment settings and experimental results are shown in [Sect. 11.3](#). We conclude the paper in [Sect. 11.4](#).

## 11.2 Measurement and Analysis Method of Bottleneck Links

### 11.2.1 Basic Idea

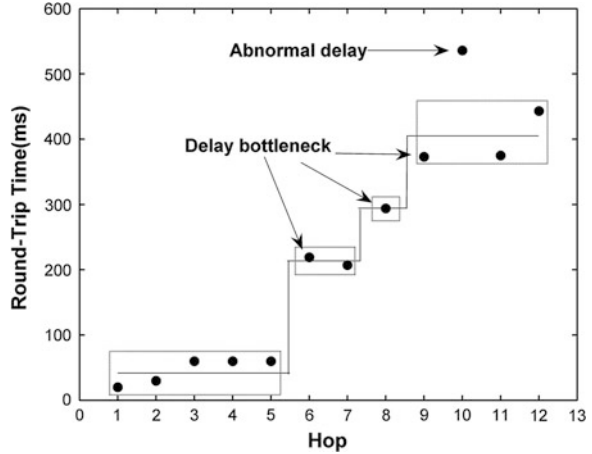
An end-to-end path is composed of many hops normally, as shown in [Fig. 11.1](#).

The source, the destination and intermediate routers on the path are called network nodes, denoted as  $a_i$ ,  $i = 0, 1 \dots n$ . The links between adjacent network nodes are denoted as  $l_i$ ,  $i = 1, 2 \dots n$ . Because of route changing, different nodes may appear in the  $i$ -th hop, leading to different values of  $a_i$ . The end-to-end path is denoted as  $(a_0, a_1 \dots a_n)$ . The delay from the source  $a_0$  to the  $i$ -th hop is denoted as  $D_i$  and  $d_i$  is the delay of link  $l_i$ ,  $i = 1, 2 \dots n$ . Thus, we have  $D_0 = 0$ ,  $d_i = D_i - D_{i-1}$ ,  $i = 1, 2 \dots n$ .

The change of link delay in an end-to-end path exhibits a non-linear pattern usually (see [Fig. 11.2](#)). We define the delay bottleneck as the link which satisfies the following conditions:

- (1) Assume the link delay is  $d_i$ ,  $d_i > DT$ , where  $DT$  is the bottleneck threshold. In other words, the delay of a bottleneck link must be high above a threshold value;
- (2) The delay of a bottleneck link should be greater than the delay of a non-bottleneck link significantly.

**Fig. 11.2** Delay bottlenecks in an end-to-end path



### 11.2.2 Measurement of Path Delay

To measure the link delays in an end-to-end path, we design and implement a parallel active probing tool based on traceroute, called OPtraceroute. It is based on OpenMP [15] parallel computing package and SOCKET networking library. Figure 11.3 describes the details of OPtraceroute.

Comparing to traditional *traceroute*, the characteristics of OPtraceroute are as following:

- (1) OPtraceroute parallelly sends and receives different *TTL* ICMP packets from intermediate routers of end-to-end path based on OpenMP, it is not like traditional *traceroute* which sends the next *TTL* ICMP packets until it receives the respond of the last. In OPtraceroute, each group of the ICMP probing packet takes up one thread. According to the CPU performance of the source,

---

**Input:** the IP address of destination

**Output:**  $(a_0, a_1 \dots a_n)$  and  $D_i, i = 1, 2 \dots n$

---

- 1 set the value of OpenMP environment variable *NUM\_THREADS*
  - 2 **for**  $TTL = j, j=1, 2, \dots, 32$  **do**
  - 3 parallelly send ICMP\_ECHO packet with  $TTL = k, j \cdot k \cdot (NUM\_THREADS + j)$ ;
  - 4 parallelly receive ICMP\_REPLY packet from  $a_k$ , record the IP address of  $a_k$  and *RTT* as:  $2 \times D_k$ ;
  - 5 **if** the time for receiving ICMP\_REPLY packet from  $a_k$  is out, **then** repeatedly send ICMP\_ECHO packet with  $TTL = k$ , **until** repeat time exceeds 3; **if** repeat time exceeds 3, **then**  $a_k$  is unreachable;
  - 6 **if** the IP address of  $a_k$  equal to the destination IP address **then break**
  - 7 stop probing and output  $(a_0, a_1 \dots a_n), D_i, i = 1, 2 \dots n$
- 

**Fig. 11.3** OPtraceroute process

the number of threads supported is also different, the maximum threads as:  $NUM\_THREADS \times \text{the numbers of destinations}$ . This method can effectively shorten detection period.

- (2) The traditional *traceroute* with the same *TTL* detects 3 times. OPtraceroute detects only once, unless timeout for the response of intermediate route, reduces duplication detection.
- (3) OPtraceroute could not use UDP packet, but ICMP, and save the result to text for analysis.

## 11.2.3 Method to Identify Multiple Bottleneck Links

### 11.2.3.1 WARD Data Clustering Method

Clustering analysis is a mathematical method to study and analyze the relationship and regularity of data. The WARD data method is a clustering analysis method based on the concept of *distance* which measures a kind of fuzzy similarity relation. The algorithm is described as follows [13].

The set  $X$  which includes  $n$  samples is divided into  $k$  categories:  $G_1, G_2, \dots, G_m$ ,  $t = k$ ;  $X_i^{(m)}$  is the  $i$ -th element of  $G_m$ , the number of the elements of  $G_m$  is  $n_m$ ,  $i = 1, 2, \dots, n_m$ ;  $\bar{X}^{(m)}$  is the average value of the elements in  $G_m$ ; the sum of square deviation of the elements in  $G_m$  is:

$$S_m = \sum_{i=1}^{n_i} \left( X_i^{(m)} - \bar{X}^{(m)} \right)' \left( X_i^{(m)} - \bar{X}^{(m)} \right) \quad (11.1)$$

Let  $G_{pq}$  be the combination of  $G_p$  and  $G_q$  and their sums of square deviation are  $S_{pq}$ ,  $S_p$  and  $S_q$ , respectively. We define the *distance* of  $G_p$  and  $G_q$  as:

$$D_{pq}^2 = S_{pq} - S_p - S_q = \frac{n_p n_q}{n_{p+q}} \left( \bar{X}^{(p)} - \bar{X}^{(q)} \right)' \left( \bar{X}^{(p)} - \bar{X}^{(q)} \right) \quad (11.2)$$

From (11.2), we get the *distance* matrix as:

$$R^{(k)} = \begin{pmatrix} D_{12}^2 & D_{13}^2 & \cdots & D_{1k}^2 \\ 0 & D_{23}^2 & \cdots & D_{2k}^2 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & D_{(k-1)k}^2 \end{pmatrix} \quad (11.3)$$

In the matrix  $R^{(k)}$ , we can find the minimum *distance*  $D_{xy}^2$ , and then we merge  $G_x$  and  $G_y$  to get  $G_{xy}$ .  $R^{(k)}$  would be reduced to a  $(k-1)$ -order matrix. The above procedure is repeated until the required order is reached.

### 11.2.3.2 Identify Bottleneck Links Using WARD Method

We analyze the data obtained through OPtraceroute using the WARD method and identify multiple bottleneck links in end-to-end paths. The specific procedure is stated as follows:

**Step 1: Data pre-processing.** Ideally, the delays of links in an end-to-end path should grow monotonically. Formally, we have  $d_1 < d_2 < \dots < d_n$  where  $n$  is the hop number. But in real measurement data, there may be  $d_i < d_j$  when  $i > j$ . Thus, we need to correct such data to make the delays of links to grow monotonically and the procedure is illustrated as follows:

- (1) Take the average of  $d_i$  for 10 continuous measurements,  $i = 1, 2, \dots, n$ . The end-to-end path ( $a_0, a_1 \dots a_n$ ) of 10 measurements must keep the same. If the end-to-end path in the  $m$ -th measurement changes, we take the average of  $d_i$  for the previous  $m - 1$  measurements,  $1 < m < 10$ .
- (2) If  $d_{i+1} < d_i$ ,  $i = 1, 2, \dots, n-1$ , then let  $d_i = d_{i+1}$  so that the delays of links grow monotonically.

**Step 2: Identify bottleneck links.** We cluster the corrected data using the WARD method until the maximal delay value  $d_{max}$  is clustered into one group. Then the links in the group are considered as bottleneck links.

## 11.2.4 Analysis of Bottleneck Links

### 11.2.4.1 Relationship of Bottleneck Links and Autonomous System

After identifying multiple bottleneck links, an interesting question is to know how the bottleneck links distribute in the Internet. Specifically, we hope to know whether the bottleneck links are inter-domain links or not. By analyzing the ratio of inter-AS and intra-AS bottleneck links, we can find whether the performance of the Internet is mainly affected by links between ASes.

Firstly, we get the mapping of IP addresses to AS numbers from the data of Routeviews [16]. Then we know a bottleneck link belongs to the category of inter-AS or intra-AS links by comparing the AS numbers of two end-points of the link. Similar to the approach in [4], we classify the link  $l_i$  into one of the following three categories:

- (1) *Inter0-AS link.* The link  $l_i$  is an inter0-AS link if the two end-points of the link do not belong to the same AS.
- (2) *Inter1-AS link.* The link  $l_i$  is an inter1-AS link if both two end-points of the link belong to the same AS and it is adjacent to an inter0-AS link. In this case,  $l_i$  appears to be one hop away from the link where AS number changes, but it might be an intercontinental link or an international link.
- (3) *Intra-AS link.* Other links which end-points belong to the same ASes are intra-AS links.

### 11.2.4.2 Relationship of Bottleneck Links and Geographical Location

The delay of end-to-end path is composed of transmission delay, propagation delay, processing delay and queuing delay [6]. Wherein, propagation delay is related to the cable length and the velocity of electrical signals. We hope to analyze the influence of propagation delay on the delay of end-to-end path through the relationship of bottleneck links and geographical locations.

We use the data set GeoIPCountryWhois [17] to map IP addresses to geographical locations. By comparing the country code of end-points of bottleneck links, we obtain the relationship of bottleneck links and geographical locations, and then analyze the distribution characteristics of bottleneck links in geographical space.

## 11.3 Experiments and Analysis

### 11.3.1 Data Collection

In the experiments, we run OPtraceroute on a host in Tianjing to measure the paths of destinations selected from 10 different domains, as described in Table 11.1. We probe each of them every 30 s for 14 days from August 13, 2012. The parameter *NUM\_THREADS* is set to 8.

The delay threshold *DT* is set as 50 ms (micro-seconds). When the delay from the source to the destination  $D_n$  is less than *DT*, there are no bottleneck links in the end-to-end path.

We try to collect 381540 sets of probing data for 10 destinations in the experiments and 350433 effective data sets are obtained based on the response of destinations. The destinations and intermediate hops locate in 30 ASes, 16 countries or regions.

**Table 11.1** Destination IP addresses

Id	Location	IP address
1	Abuja, Nigeria	41.78.83.94
2	Seoul, South Korea	58.229.14.62
3	Israel	62.219.189.149
4	Barbados	69.80.55.6
5	Madrid, Spain	80.38.204.121
6	Russian Federation	87.226.230.250
7	New York, USA	216.213.16.22
8	Marseille, France	217.16.0.2
9	Taiwan, China	220.128.6.126
10	Lhasa, China	219.151.32.2

**Table 11.2** Average number of bottleneck links

Id	1	2	3	4	5	6	7	8	9	10
Ave. number	2.7	0.58	1.85	2	2.29	1.92	1.42	1.87	0.52	0.1

Only 23.2 % probings of Marseille are successful. From 2012/8/23 13:29:20 to 2012/8/26 15:6:20, the destination IP address can be reached. But in other time, the last-hop which can be probed is always 86.79.0.22. We guess that it fails due to certain routing events.

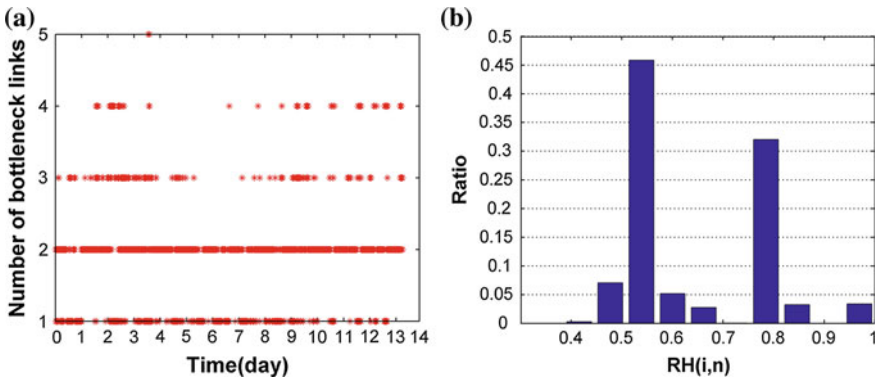
### 11.3.2 Result Analysis

#### 11.3.2.1 Identification of Multiple Bottleneck Links

Using the proposed method, we get identification results of bottleneck links. The average numbers of bottleneck links are shown in Table 11.2. From the table, we can see that the average number takes the lowest value for Seoul (Id = 2), Taiwan (Id = 9) and Lhasa (Id = 10), while Abuja (Id = 1), Barbados (Id = 4) and Madrid (Id = 5) have the highest average bottleneck link numbers. Figure 11.4a shows the variation of bottleneck link numbers for the location 3 (Israel). It shows that the number varies between 1 and 4 mainly.

#### 11.3.2.2 Distribution of Bottleneck Links in End-to-End Paths

By checking the entrance and exit IP addresses of bottleneck links, we find that the bottleneck links keep constant in most measurements. Take the IP address



**Fig. 11.4** Variation of bottleneck link numbers and frequency histogram of reference locations for location 3 (Israel)

**Table 11.3** Statistics of the main bottleneck links for location 3 (Israel)

Link Id	Entrance IP address	Exit IP address	Occurrence number (ratio)
1	202.97.53.218	202.97.51.186	3292 (33 %)
2	77.67.71.101	89.149.180.226	1262 (12 %)
3	202.97.53.214	202.97.52.186	984 (9.7 %)
4	218.30.54.70	89.149.182.226	900 (8.9 %)
5	77.67.71.101	89.149.186.42	775 (7.7 %)
6	202.97.53.218	202.97.51.62	641 (6.3 %)

62.219.189.149 as an example. In the measurements, 23 links have been classified as bottleneck links. The sum of occurrence of bottleneck links is 10102. The first 6 bottleneck links in Table 11.3 appear for 7854 times. In other words, about 26 % bottleneck links take the ratio of 77 % of all occurrence.

To study the distribution of bottleneck links in end-to-end paths, we define the *reference location* of a bottleneck link as follows. If a bottleneck link is the  $i$ -th hop in an end-to-end path of length  $n$ , then its reference location is computed as  $i/n$ , denoted by  $RH(i, n)$ . Obviously, the value of  $RH(i, n)$  is in  $[0, 1]$ . The larger value of  $RH(i, n)$ , the closer the bottleneck link is to the destination. By calculating the reference locations of all bottleneck links, we find that the bottleneck links mainly locate at few hops which are in the middle of end-to-end paths or near the destinations. Take the IP address 62.219.189.149 (location 3, Israel) as an example. Figure 11.4b shows the frequency histogram of reference locations of bottleneck links for the destination. The main bottleneck links for location 3 are shown in Table 11.3. From the table, we can see that the link (202.97.53.218, 202.97.51.186) appears as a bottleneck link mostly.

### 11.3.2.3 Distribution of Bottleneck Links Among ASes

Using the mapping from IP addresses to AS numbers, we examine the distribution of bottleneck links in the inter-domain routing systems. The top sub-figure in Fig. 11.5 shows the distribution of bottleneck links in measurements of the 10 destinations across the three categories in Sect. 11.2.4.1. It shows that bottleneck links in the measurements of New York and Marseille occur mostly as intra-AS links although we give a conservative definition of intra-AS link. We further relax the definition of inter1-AS as *two hops away* from the link where AS number changes, as shown in the bottom sub-figure. We can see that there is a great change in  $Id$  7, 8 and all of the bottleneck links rarely occur as the type of inter1-AS link. We guess that bottleneck links often occur as inter-AS links or close to inter-AS links.

Through analyzing identification results in the experiments, except the measurements of Seoul, Taiwan and Lhasa, there is one bottleneck link in AS 4134 which appears at least in other 7 measurements. It locates at Beijing and often occurs at *one or two hop away* from the link where AS number changes. By more



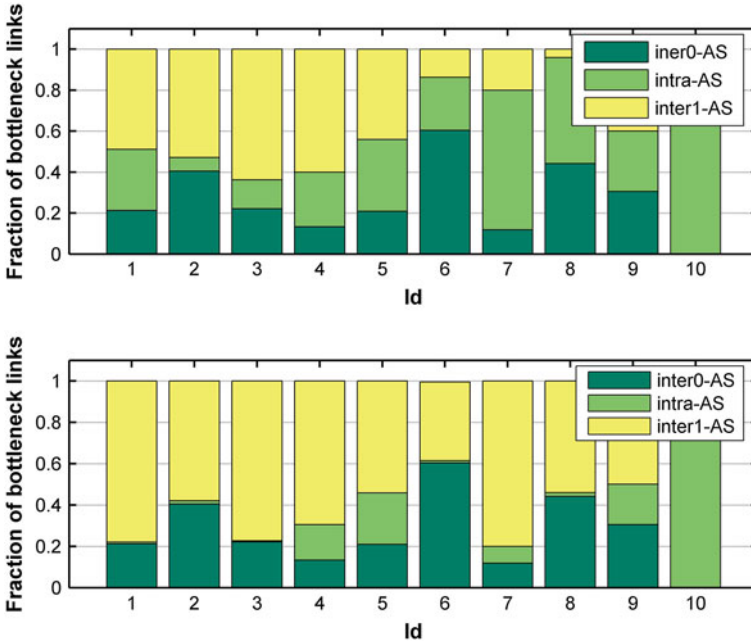


Fig. 11.5 Distribution of bottleneck links among ASes

accurate inquiry to these bottleneck links [18], we find that they are the international exits of AS 4134 backbone.

In the measurements of Seoul and Taiwan, the international exits of AS 4134 are mainly at Guangzhou and Shanghai. There are no end-to-end paths passing through Beijing. The bottleneck links are almost outside of AS 4134.

For Lhasa, links on end-to-end paths locate in the same AS 4134, and the delay from the source to the destination is almost always less than 50 ms. It seems that the network performance inside AS 4134 is good in our measurements.

In summary, the international exits of AS 4134 at Beijing are performance bottleneck when accessing foreign Internet sites. The reason may be due to iBGP routing strategies or a large amount of data traverse through these international exits.

## 11.4 Conclusion

In this paper, we have designed a parallel active measurement framework to actively measure the path delays of many destinations simultaneously and an algorithm to identify multiple bottleneck links has been proposed. The experimental results show that the proposed approach can fast, accurately obtain path delays and locate multiple bottleneck links. Meanwhile, we have found that bottleneck links are mainly few constant links. The number of intra-domain

bottleneck links takes a large portion in most cases, which hints that the performance of end-to-end paths may be greatly influenced by iBGP routing. The intercontinental links in an anonymous system incline to be bottleneck links in end-to-end paths.

Due to the limitation of the experimental data, we could not fully investigate the relationship of bottleneck links with geographical location and ASes. Many issues require further study by increasing the experimental scale. They include studying the stability of bottleneck links, the impact of routing change on bottleneck links. We also hope to improve the data clustering method in identifying multiple bottleneck links, for example, using better data clustering method for different types of paths.

**Acknowledgments** This work was supported in part by the National Natural Science Foundation of China under grant No.61070199, 61103189, 61003301 and 61100223, and 863 High-Tech. Program of China No. 2011AA01A103.

## References

1. Jacobson V (1997) Pathchar—a tool to infer characteristics of internet paths. *Math Sci Res Inst*
2. Hu N, Steenkiste P (2003) Evaluation and characterization of available bandwidth probing techniques. *IEEE JSAC (Special Issue in Internet and WWW Measurement, Mapping, and Modeling)* 21(6):879–894
3. Akella A, Seshan S, Shaikh A (2003) An empirical evaluation of wide-area internet bottlenecks. In: *IMC '03 proceedings of the 3rd ACM SIGCOMM conference on internet measurement*, pp 101–114
4. Hu N, Li L, Mao Z, Steenkiste P, Wang J (2004) Locating internet bottlenecks: algorithms, measurements, and implications. In: *SIGCOMM '04 proceedings of the 2004 conference*, pp 41–54
5. He X, Papadopoulos C, Heidemann J, Mitra U, Riaz U (2009) Remote detection of bottleneck links using spectral and statistical methods. *Comput Netw* 53:279–298
6. Hu X, He J, Shi H (2009) Inference of network bottleneck link based on end-to-end loss measurements. *J Beijing Univ Posts Telecommun* 32:93–96
7. Choi B-Y, Moon S, Zhang Z-L, Papagiannaki K, Diot C (2004) Analysis of point-to-point packet delay in an operational network. In: *Proceedings of INFOCOM 2004*, pp 1797–1807
8. Govindan R, Paxson V (2002) Estimating router ICMP generation delays. In: *Passive and active measurement workshop (PAM)*
9. Almes G, Kalidindi S, Zekauskas M (1999) A round-trip delay metric for IPPM. In: *RFC 2681*
10. Bi J, Wu Q, Li Z (2003) Measurement and analysis of internet delay bottlenecks. *Chin J Comput* 26:406–416
11. Li C, Zhao H, Zhang X, Yuan S (2008) Research on the characteristic behavior of internet dominant delay. *Acta Electronica Sinica* 36:1063–1067
12. CAIDA <http://www.caida.org>
13. Yuan Z, Song S (2009) *Multivariate statistical analysis*. Science Press, ShanXi, China, pp 278–293
14. He X (2008) *Multivariate statistical analysis*. China Renmin University of Press, Beijing
15. Chandra R, Menon R, Dagum L, Kohr D, Maydan D, McDonald J (2000) *Parallel programming in OpenMP*. Academic Press, San Diego
16. Routeviews <http://www.routeviews.org>
17. GeoIPCountryWhois <http://www.hostip.info>
18. LBASE <http://www.lbase.net/>

# Chapter 12

## Hand Segmentation Based on Skin Tone and Motion Detection with Complex Backgrounds

Xintao Li, Can Tang, Chun Gong, Sheng Cheng and Jianwei Zhang

**Abstract** Hand Segmentation is the first problem need to be solved in hand recognition system. Currently, most hand gesture recognition system is based on simple background, or requests the recognizer on glove in special color, which gives human–computer interaction some restrictions. This paper researches the gesture segmentation technology based on complex backgrounds, and gives a method combined with skin tone detection and motion detection. By experiments on the images captured by home security robot, this method can get accurate hand segmentation of all the images. This paper lays the foundation of gesture recognition on the home security robots.

**Keywords** Complex background · Hand segmentation · Skin color · Motion detection · Home security robot

### 12.1 Introduction

Gesture is a natural and intuitive interpersonal communication mode, therefore, in the field of human–computer interaction, Gesture recognition is the hot research topic, gesture recognition based on sequences (images) is the indispensable key technology of the new generation of human–computer interaction. Realize gesture recognition system need to solve the three important problems [1]: gesture segmentation, gesture analysis and gesture recognition. With the influence of complex of background and environment light, in the gesture recognition method based on monocular vision, how to division out gesture region is always a difficulty, many researchers used the method of limiting the gesture image, for example, use the

---

X. Li (✉) · C. Tang · C. Gong · S. Cheng · J. Zhang  
Intelligent Robot Engineering Lab of Kunshan ITRI, 6F ITRI Building, No.1699, Weicheng  
South Road, Kunshan, Jiangsu, China  
e-mail: lixt@ksitri.com

pure black or white wall, simplified background by dressed in the dark black clothing, or require people to wear special color gloves for outstanding hands area, etc. however, These methods increased the limitation of human–computer interaction, destroy the system availability and user-friendliness.

This paper mainly studies the gesture segmentation method within complex background, it segments hand Based on Skin Tone and Motion Detection. First, segment skin areas from complex background use skin color model. Then, get the moving regions by motion detection and filter the still skin areas in the background by mask the skin areas and moving regions. Last, get the accurate hand area by masking the moving hand areas and skin areas.

## 12.2 Skin Segmentation

The purpose of skin segmentation [2] is to separate the skin areas from the complex background, skin segmentation need to select appropriate color space and establish skin model. This paper uses the Gaussian skin model based on  $YCbCr$  color space.

$YCbCr$  color space [3] can separate the luminance and chrominance of the image,  $Y$  component indicate the brightness of the pixel,  $Cb$  and  $Cr$  components called chrominance,  $Cb$  indicates blue component,  $Cr$  indicates red component, Color in this color space can be gathered in a very small range.  $YCbCr$  color space can full disclosure skin of body, and can maximum eliminate the influence of brightness, so reduce the number of dimensions of color space and reduce the computational complexity. We usually need to convert the  $RGB$  color space to  $YCbCr$  color space, the transformation formula as follows:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & 0.331 & 0.500 \\ 0.500 & -0.419 & 0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (12.1)$$

Gaussian model [4] mainly use the principles of statistics, it believes random samples which conform to the normal distribution also meet Gaussian distribution such as skin color. The mathematical expression of Gaussian distribution is simple, intuitive, and is a normal model which research deeper in principle of Statistics. Gaussian model constitute a continuous data information by calculating the probability of pixel value and get a probability graph of skin color, then complete color confirmation by the probability of skin color. Gaussian can express as  $N(m, C)$ ,  $m$  is mean value,  $C$  is covariance matrix.

$$m = E\{x\}, \quad x = (Cr, Cb)^T \quad (12.2)$$

$$C = E\{(x - m)(x - m)^T\} \quad (12.3)$$

By the experimental statistics, Mean and covariance matrix respectively as:

$$m = (150.3179, 117.1057)^T \quad (12.4)$$

$$C = \begin{bmatrix} 250.2594 & 18.2077 \\ 18.2077 & 149.6103 \end{bmatrix} \quad (12.5)$$

By the Gaussian skin color model establish in advance, the probability of any pixel belongs to the skin can be calculated by the following formula:

$$P(Cr, Cb) = \exp[-0.5(x - m)^T C^{-1}(x - m)] \quad (12.6)$$

Compute the skin color likelihood of all the pixels in the detected image, and get the maximum of skin color likelihood, then use the skin color likelihood of all the pixel divide the maximum skin color likelihood, we get the probability of the pixel belongs to the skin color. The image composed by skin color probability of all the pixel is called color likelihood image, in the color likelihood image, we set a threshold, when the pixel value greater than the threshold, we can confirm the pixel is skin pixel, then we can get the segmentation skin image. At last, corrode and dilate skin color detection result image, some skin like small areas can be eliminated. The results are shown as Fig. 12.1.

### 12.3 Motion Detection

The purpose of motion detection is to extract the changed area in the sequence images from the background. The effective segmentation of the moving regions is essential to the later processing of target classification, tracking and behavior understanding. However, because of the dynamic changes of background image, for example, influence by the weather, illumination and shadow, make the motion detection to be a very difficult work. Commonly used motion detection methods are Background Subtraction [5, 6], Temporal Difference [7] and Optical Flow [8].

In the hands waved process, hands is a motion area, therefore, we can eliminate the disturbance of color like regions in the static background by motion information. Based on the efficiency of algorithm consideration, this paper uses the method of Temporal Difference for motion detection. Temporal Difference (also



**Fig. 12.1** Skin segmentation **a** original image, **b** result of skin segmentation, **c** corrosion and expansion results

called Adjacent frame Difference) method extract the moving area by temporal difference based on pixel in continuous image sequence and threshold. Temporal Difference has strong adaptability for dynamic environment. The shortcoming of this method is can't detect the overlap part of the moving object, caused incomplete of the moving object, and produce empty in the internal of the moving object. In order to solve the problems, this paper selects the discontinuous and frames which have obvious movement for difference, as shown in Fig. 12.2.

From the results above, we can see that motion detection not only detected the moving hand region, but also detected the body and head movement, these movement are not we need, so in order to exactly segment moving hand region, we need to further remove the useless areas.

## 12.4 Hand Segmentation

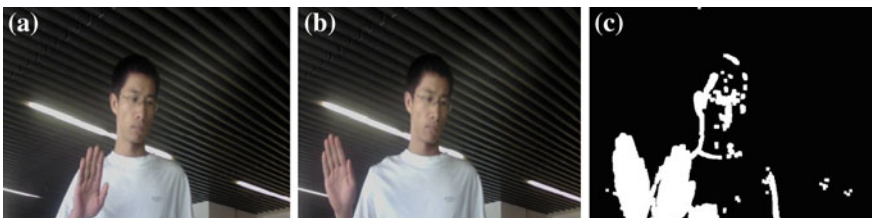
To perform the mission of hand segmentation three phases are introduced.

### 12.4.1 Skin Color Mask

Mask the motion detect result and skin color detect result, here we use and operation, by this step, we can effectively remove the moving non-skin regions, include body and other moving objects in the image, get the moving skin color regions. The results of this step are shown in Figs. 12.3 and 12.4.

### 12.4.2 Motion Mask

From the results of Sect. 12.4.1, we can see that, the results of masked motion regions and skin color regions contain the moving face regions, so we need to further eliminate these regions. By analyzing the results of the first step, we get the



**Fig. 12.2** Motion detection **a** frame 1, **b** frame 10, **c** motion detection results



**Fig. 12.3** Skin color mask (1) **a** motion detection result, **b** skin detection result (1), **c** mask result (1)



**Fig. 12.4** Skin color mask (2) **a** motion detection result, **b** skin detection result (2) **c** mask result (2)

conclusion that again masks the moving skin regions can eliminate the face skin regions. The result of this step is shown as Fig. 12.5.

### 12.4.3 Hand Region Extraction

From the results of Sect. 12.4.2, we can see that, skin like regions in the background and face regions are completely eliminated. But both of the two hands are saved. Obviously, again masks the results of motion mask and skin detection, we can get the signal hand region in the current image. The results of this step are shown as Figs. 12.6 and 12.7.



**Fig. 12.5** Motion mask **a** skin color mask result (1), **b** skin color mask result (2), **c** motion mask result



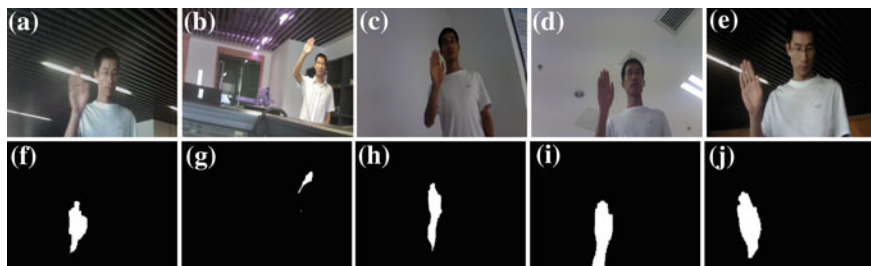
**Fig. 12.6** Hand region extraction (1) **a** skin color detect result (1), **b** motion mask result, **c** hand segmentation result (1)



**Fig. 12.7** Hand region extraction (2) **a** skin color detect result (2), **b** motion mask result, **c** hand segmentation result (2)

## 12.5 Experiments

In order to check the validity of the method proposed in this paper, we use the surveillance camera in the home security robot shooting different environment of videos, select different video frames for experiments, at the same time, use this method to hand wave direction recognition of the home security robot. From the results, this method achieve recognition rate of 90 %, which greatly improving the interactive performance of the home security robot. Some hand region segmentation results are shown as Fig. 12.8a–e are original images, f–j are corresponding hand region segmentation results.



**Fig. 12.8** Hand region segmentation results



## 12.6 Conclusion

This paper aims at the difficulties of hand region segmentation with complex background, uses a method combined with Skin Tone and Motion Detection, realize accurate hand region segmentation from coarse to fine. First, use the Gaussian skin color model based on  $YCbCr$  color space to detect skin region. Then, detect motion from two discontinuous image frames, eliminate non-skin moving regions by skin mask and eliminate face skin regions by motion mask. In the end, again mask the results of motion mask and skin regions, accurately segment out the hand region in the current image. Through the results of the experiments, this method can better segment hand regions under different complex backgrounds, achieve accurate hand wave direction recognition, greatly enhance interactive performance of the home security robot.

## References

1. Chen Y, Zhang Y (2009) Research on human-robot interaction technique based on hand gesture recognition. *Robot* 31(4):351–356 (in Chinese)
2. Chen D, Liu Z (2006) A survey of skin color detection. *Chin J Comput* 29(2):194–203 (in Chinese)
3. Zheng N (1998) *Computer vision and pattern recognition*. National Defence Industry Press, Beijing
4. Xu Z, Zhu M (2007) Color-based skin detection: a survey. *J Image Graph* 12(3):377–388 (in Chinese)
5. McKenna SJ, Jabri S, Duric Z (2000) Tracking groups of people. *Comput Vis Image Underst* 80(1):42–56
6. Haritaoglu I, Harwook D (2000) W4: real-time surveillance of people and their activities. *IEEE Trans Pattern Mach Intell* 22(8):809–830
7. Lipton A, Fujiyoshi H, Patil R (1998) Moving target classification and tracking from real-time video. In: *Proceedings of IEEE workshop an application of computer vision*, Princeton, NJ, pp 8–14
8. Meyer D, Denzler J, Niemann H (1997) Model based extraction of articulated objects in image sequence for gait analysis. In: *IEEE international conference on image processing*, vol 3, pp 78–81

# Chapter 13

## Local Reconstruction and Dissimilarity Preserving Semi-Supervised Dimensionality Reduction

Feng Li, Zhengqun Wang, Zhongxia Zhou and Wei Xue

**Abstract** In this paper, a semi-supervised dimensionality reduction algorithm for feature extraction, named LRDPSSDR, is proposed by combining local reconstruction with dissimilarity preserving. It focuses on local and global structure based on labeled and unlabeled samples in learning process. It sets the edge weights of adjacency graph by minimizing the local reconstruction error and preserves local geometric structure of samples. Besides, the dissimilarity between samples is represented by maximizing global scatter matrix so that the global manifold structure can be preserved well. Comprehensive comparison and extensive experiments demonstrate the effectiveness of LRDPSSDR.

**Keywords** Local reconstruction · Dissimilarity preserving · Semi-supervised dimensionality reduction · Face recognition

### 13.1 Introduction

Face recognition has become one of the most challenging problems in the application of pattern recognition. Face image is a high dimension vector, so numerous dimension reduction techniques have been proposed over the past few decades [1], in Principal component analysis (PCA) [2] and Linear discriminant analysis (LDA) [3] are widely used. Both PCA and LDA assume feature space lie on a linearly embedded manifold and aim at preserving global structure. However, many researches have shown that the face images possibly reside on a nonlinear submanifold [4]. When using PCA and LDA for dimensionality reduction, they will fail to discover the intrinsic dimension of image space.

---

F. Li (✉) · Z. Wang · Z. Zhou · W. Xue  
Department of Automation, College of Information and Engineering,  
Yangzhou University, Yangzhou 225127, China  
e-mail: yzulifeng@126.com

By contrast, manifold learning considers the local information of samples, aiming to directly discover the globally nonlinear structure. The most important manifold learning algorithms include isometric feature mapping (Isomap) [5], locally linear embedding (LLE) [6], and Laplacian eigenmap (LE) [7]. Though these methods are appropriate for representation of nonlinear structure, they are implemented restrictedly on training samples and can not show explicit maps on new testing samples in recognition. Therefore, locality preserving projection (LPP) [8] is proposed, but it only can focus on the local information of training samples. To remedy this deficiency, unsupervised discriminant analysis (UDP) [9] is introduced, which can consider the local structure as well as global structure of samples. However, it only uses unlabeled data. Therefore, semi-supervised methods to deal with insufficient labeled data could be learned. It can be directly applied in the whole input space, while the out-of-sample problem can be effectively solved.

This paper simultaneously investigates two issues. First, How to extract the effective discriminant feature by using labeled samples? Second, How to minimize local scatter matrix and maximize global scatter matrix simultaneity, to preserve the local and global structure information?

The rest of this paper is organized as follows. In Sect. 13.2, we give the details of LRDPSDR. The experimental results based on two commonly used face databases demonstrate the effectiveness and robustness of proposed method in Sect. 13.3. Finally, the conclusions are summarized in Sect. 13.4.

## 13.2 Locally Reconstruction and Dissimilarity Preserving Semi-Supervised Dimensionality Reduction

### 13.2.1 Local Reconstruction Error

Suppose  $\mathbf{X} = [x_1, x_2, \dots, x_l, x_{l+1}, \dots, x_{l+u}]$  be a set of training samples that include  $l$  labeled samples and  $u$  unlabeled samples, belonging to  $c$  classes. For each sample, we find its  $k$  nearest neighbors  $N_k(x_i)$  from  $\mathbf{X}$ , where  $N_k(x_i)$  is the index set of the  $k$  nearest neighbors of  $x_i$ . To preserve local structure, design following objective function:

$$J(a) = \sum_{i=1}^{l+u} \left\| x_i - \sum_{j: x_j \in N_k(x_i)} C_{ij} x_j \right\|^2 \quad (13.1)$$

where  $\|\bullet\|$  denotes the Euclidean norm, with two constraints:

- (1)  $\sum_j C_{ij} = 1, i = 1, 2, \dots, l + u$
- (2)  $C_{ij} = 0$ , if  $x_j$  does not belong to the set of  $k$  nearest neighbors of  $x_i$ .

Given samples in low-dimensional linear embedding space as follows:

$$\begin{aligned}
\sum_{i=1}^{l+u} \left\| y_i - \sum_{j:x_j \in N_k(x_i)} C_{ij} y_j \right\|^2 &= \sum_{i=1}^{l+u} \left( \mathbf{w}^T x_i - \sum_{j:x_j \in N_k(x_i)} C_{ij} \mathbf{w}^T x_j \right)^T \left( \mathbf{w}^T x_i - \sum_{j:x_j \in N_k(x_i)} C_{ij} \mathbf{w}^T x_j \right) \\
&= \sum_{i=1}^{l+u} \left( \mathbf{w}^T \left( x_i - \sum_{j:x_j \in N_k(x_i)} C_{ij} x_j \right) \right)^T \left( \mathbf{w}^T \left( x_i - \sum_{j:x_j \in N_k(x_i)} C_{ij} x_j \right) \right) \\
&= \text{tr} \left\{ \mathbf{W}^T \left[ \sum_{i=1}^{l+u} \left( x_i - \sum_{j:x_j \in N_k(x_i)} C_{ij} x_j \right) \left( x_i - \sum_{j:x_j \in N_k(x_i)} C_{ij} x_j \right)^T \right] \mathbf{W} \right\} \\
&= \text{tr}(\mathbf{W}^T \mathbf{S}_L \mathbf{W})
\end{aligned} \tag{13.2}$$

where  $\text{tr}(\bullet)$  is the notation of trace. Local scatter matrix is defined as follows:

$$\mathbf{S}_L = \mathbf{X} \mathbf{M} \mathbf{X}^T \tag{13.3}$$

where  $\mathbf{M} = (\mathbf{I} - \mathbf{C})^T (\mathbf{I} - \mathbf{C})$ ,  $\mathbf{I}$  is an  $(l+u) \times (l+u)$  identity matrix, and matrix  $\mathbf{C}$  is calculated by LLE [6] algorithm.

With above two constraints, the local reconstruction weight is invariant to rotation, rescaling, and translation. So, it can preserve local structure of samples.

### 13.2.2 Dissimilarity Preserving

Let  $\mathbf{X} = [x_1, x_2, \dots, x_l, x_{l+1}, \dots, x_{l+u}]$  be a matrix and each column is a training samples, belonging to  $c$  classes. Represent any  $x_i$  in term of its projection sequence  $G(x_i)$ , coefficient vector  $W(x_i, G(x_i))$ , and residue  $R(x_i, G(x_i))$ . Suppose  $G(x_i) = [g_{i1}, g_{i2}, \dots, g_{id}]$  are the eigenvectors corresponding to the first  $d$  largest eigenvalues in PCA. Thus, each image can be represented as a linear combination:

$$\hat{x}_i = w_{i1} g_{i1} + w_{i2} g_{i2} + \dots + w_{id} g_{id} \tag{13.4}$$

where  $\hat{x}_i$  denotes the approximation of  $x_i$ , and the residue  $R(x_i, G(x_i))$  is given by

$$R(x_i, G(x_i)) = x_i - (w_{i1} g_{i1} + w_{i2} g_{i2} + \dots + w_{id} g_{id}) = x_i - \sum_{j=1}^d w_{ij} g_{ij} \tag{13.5}$$

When  $x_1$  is projected on projection sequence  $G(x_2)$  of  $x_2$ , noting the corresponding coefficient vector  $W(x_1, G(x_2))$  and residue  $R(x_1, G(x_2))$ . Based on above factor, we can design the dissimilarity measure as follows:

$$\varphi(x_1, x_2) = \sqrt{\xi D_R(x_1, x_2) + (1 - \xi) D_W(x_1, x_2)} \tag{13.6}$$

where  $\zeta \in [0, 1]$  indicates the relative importance of the residue and the corresponding coefficients, when both are projected onto the projection sequence  $G(x_2)$  of  $x_2$ , we have

$$D_R(x_1, x_2) = \|R(x_1, G(x_2)) - R(x_2, G(x_2))\|^2 \quad (13.7)$$

where  $D_R(x_1, x_2)$  is the difference between the residues of  $x_1$  and  $x_2$ , and  $D_W(x_1, x_2)$  compares their corresponding coefficients, which is represented as:

$$D_W(x_1, x_2) = \|W(x_1, G(x_2)) - W(x_2, G(x_2))\|^2 \quad (13.8)$$

Define the objective function of dissimilarity preserving as follows:

$$\sum_{i=1}^{l+u} \sum_{j=1}^{l+u} (y_i - y_j)^2 \mathbf{H}_{ij} \quad (13.9)$$

where  $\mathbf{H}_{ij}$  is the dissimilarity weight matrix, and  $\mathbf{H}_{ij} = \varphi(x_i, x_j)$ .

$$\begin{aligned} \frac{1}{2} \sum_{i=1}^{l+u} \sum_{j=1}^{l+u} (y_i - y_j)^2 \mathbf{H}_{ij} &= \frac{1}{2} \sum_{i=1}^{l+u} \sum_{j=1}^{l+u} (\mathbf{W}^T x_i - \mathbf{W}^T x_j)^2 \mathbf{H}_{ij} = \mathbf{W}^T \frac{1}{2} \sum_{i=1}^{l+u} \sum_{j=1}^{l+u} (x_i - x_j)(x_i - x_j)^T \mathbf{H}_{ij} \mathbf{W} \\ &= \mathbf{W}^T \frac{1}{2} \left( \sum_{i=1}^{l+u} \sum_{j=1}^{l+u} x_i x_i^T \mathbf{H}_{ij} + \sum_{i=1}^{l+u} \sum_{j=1}^{l+u} x_j x_j^T \mathbf{H}_{ij} - 2 \sum_{i=1}^{l+u} \sum_{j=1}^{l+u} x_i x_j^T \mathbf{H}_{ij} \right) \mathbf{W} \\ &= \mathbf{W}^T (\mathbf{X} \mathbf{D} \mathbf{X}^T - \mathbf{X} \mathbf{H} \mathbf{X}^T) \mathbf{W} = \mathbf{W}^T \mathbf{S}_N \mathbf{W} \end{aligned} \quad (13.10)$$

Further, the global scatter matrix is defined as follows:

$$\mathbf{S}_N = \mathbf{X} \mathbf{L} \mathbf{X}^T \quad (13.11)$$

where  $\sum \mathbf{H}_{ij}$  is a diagonal matrix,  $\mathbf{L} = \mathbf{D} - \mathbf{H}$  is a Laplacian matrix.

As a result, by maximizing the global scatter matrix can make the nearby samples of the same class become as compact as possible and simultaneously the nearby samples belonging to different classes become as far as possible.

### 13.2.3 The Algorithm of LRDPSDR

According to above detail analysis, we have two scatter matrices based on labeled data and unlabeled data. Associated them with Fish criterion, a semi-supervised learning method will be derived. Design the objective function as follows:

$$\mathbf{W}_{opt} = \arg \max_{\mathbf{W}} \frac{\mathbf{W}^T (\mathbf{S}_b + \beta \mathbf{S}_N) \mathbf{W}}{\mathbf{W}^T (\mathbf{S}_t + \alpha \mathbf{S}_L) \mathbf{W}} = \arg \max_{\mathbf{W}} \frac{\mathbf{W}^T (\mathbf{S}_b + \beta \mathbf{X} \mathbf{L} \mathbf{X}^T) \mathbf{W}}{\mathbf{W}^T (\mathbf{S}_t + \alpha \mathbf{X} \mathbf{M} \mathbf{X}^T) \mathbf{W}} \quad (13.12)$$

where  $\alpha, \beta > 0$  are the regularization parameters. The optimization problem solution of formula (13.12) could obtain by following eigen-equation:

$$(\mathbf{S}_b + \beta \mathbf{X} \mathbf{L} \mathbf{X}^T) \mathbf{W} = \lambda (\mathbf{S}_t + \alpha \mathbf{X} \mathbf{M} \mathbf{X}^T) \mathbf{W} \quad (13.13)$$

Based on above discussion, the proposed algorithm LRDPSDDR is briefly stated as below:

Step 1: For the given training data set, use PCA for dimensionality reduction.

Step 2: Calculate within-class scatter matrix  $\mathbf{S}_b$  and total scatter matrix  $\mathbf{S}_t$  for labeled data.

Step 3: Construct local scatter matrix  $\mathbf{S}_L$  using formula (13.3) and global scatter matrix  $\mathbf{S}_N$  using formula (13.11).

Step 4: The optimal transformation matrix  $\mathbf{W}$  is formed by the  $d$  eigenvectors corresponding to the first  $d$  largest non-zero eigen-value of formula (13.13).

Step 5: Project the training data set onto the optimal projection vectors obtained in Step 4, and then use the nearest neighbor classifiers for classification.

## 13.3 Experiments

In this section, we investigate the performance of our algorithm for face recognition and compare it with PCA, LDA, LPP, and UDP. The KNN classifier is used. The regularization parameters are set as  $\alpha = \beta = 0.1$ ,  $\zeta = 0.5$ .

### 13.3.1 Experiment on the Yale Dataset

The database consists of 165 face images of 15 individuals. These images are taken under different lighting condition and different facial expression. All images are gray scale and normalized to a resolution of  $100 \times 80$  pixels. In the experiment, we select the first five images from each individual to form the training samples and the remaining six images as testing samples. Some typical images are shown in Fig. 13.1.

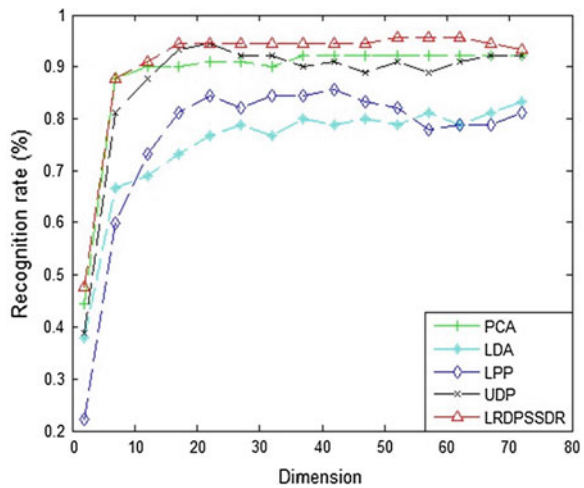
Figure 13.2 show that the recognition rate of LRDPSDDR method has significantly improvement compared to other four algorithms with the increase in the number of projection axis dimension. The projection axis dimension from 20 to 75 stages, the recognition rate tends stable. It can achieve a maximum value when projection axis dimension reach 55.

Table 13.1 gives the maximum recognition rate of five algorithms. It is not difficult to see that our method is the best according to Table 13.1 and Fig. 13.2.



**Fig. 13.1** Sample images for one individual in the YALE database

**Fig. 13.2** Comparison of recognition rate on YALE database



**Table 13.1** Comparison on top accuracy

Algorithms	PCA	LDA	LPP	UDP	LRDPSSDR
Recognition rate (%)	92.22	83.33	85.86	92.22	95.56
Dimension	35	75	40	70	55

**Table 13.2** Comparison on top accuracy

Algorithms	PCA	LDA	LPP	UDP	LRDPSSDR
Recognition rate (%)	62.98	58.81	61.19	64.52	70.71
Dimension	110	140	160	160	140



Fig. 13.3 Part of the sample images for one individual in the AR database

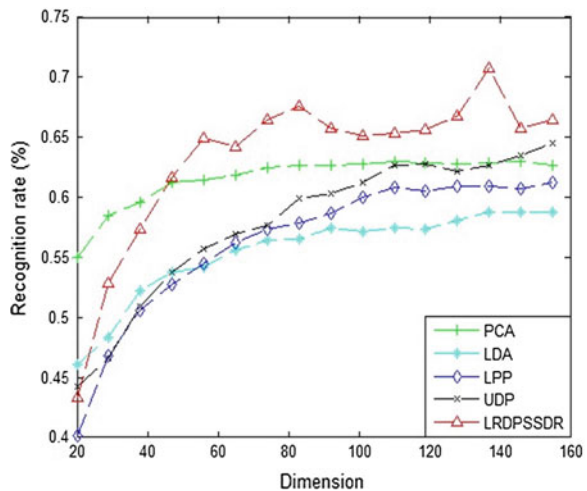
### 13.3.2 Experiment on the AR Dataset

The database consists of 3120 face images 120 individuals. These face images are captured under varying facial expressions, lighting conditions and occlusions. The size of every face image is gray scale and normalized to a resolution of  $50 \times 40$  pixels. In the experiment, we select the first seven images from each individual to form the training samples and next seven images (most are covered) as testing samples. Some typical images are shown in Fig. 13.3.

From Fig. 13.4, we can see first that the proposed LRDPSSDR method outperforms PCA, LDA, LPP and UDP, and second that our method is more robust in different lighting conditions and various facial expressions. With the projection axis dimension increase, the recognition rate raises from 30 to 140 stages. It can achieve a maximum value when projection axis dimension is 140.

Table 13.2 shows the maximum recognition rate of five algorithms. It is obvious that our method is better than other methods, so the effectiveness and robustness of LRDPSSDR is further verified.

Fig. 13.4 Comparison of recognition rate on AR database





## 13.4 Conclusion

In this paper, we present a semi-supervised learning algorithm LRDPSDD for dimension reduction, which can make use of both labeled and unlabeled data. The algorithm is realized based on both local reconstruction error and dissimilarity preserving, which not only preserves the intraclass compactness and the interclass separability, but also describes local and global structure of samples.

**Acknowledgments** We wish to thank the National Science Foundation of China under Grant No. 61175111, and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant No. 10KJB510027 for supporting this work.

## References

1. Cevikalp H, Neamtu M, Wilkes M, Barkana A (2005) Discriminative common vectors for face recognition. *IEEE Trans Pattern Anal Mach Intell* 27(1):4–13
2. Turk M, Pentland A (1991) Eigenface for recognition. *J Cogn Neurosci* 3(3):72–86
3. Zheng WS, Lai JH, Yuen PC (2009) Perturbation LDA: learning the difference between the class empirical mean and its expectation. *Pattern Recogn* 42(5):764–779
4. Chen H-T, Chang H-W, Liu T-L (2005) Local discriminant embedding and its variants. In: *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, vol 2, pp 846–853
5. Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290:2319–2323
6. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* 290:2323–2326
7. Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput* 15(6):1373–1396
8. He XF, Niyogi, P (2003) Locality preserving projections. In: *Proceedings of the 16th annual neural information processing systems conference*, Vancouver, pp 153–160
9. Yang J, Zhang D (2007) Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics. *IEEE Trans Pattern Anal Intell* 29(4):650–664

# Chapter 14

## Fusion of Gray and Grads Invariant Moments for Feather Quill Crease Recognition

Hongwei Yue, Renhuang Wang, Jinghua Zhang and Zuihong He

**Abstract** In order to overcome the non-crease misjudgment of feather quill, a novel decision fusion algorithm is proposed. An improved Radon transformation is used to extract moment invariants of gray and grads dual-mode of target region and singular value decomposition is provided here to obtain feature vectors, respectively; then creases recognition is performed according to feature vectors of the dual-mode. Finally, the final recognition result of the system is achieved by the fusion of recognition results of the dual-mode at the decision level. Experimental results show that this new method can overcome the limitations of single-modal and reduce the misjudgment of non-crease effectively.

**Keywords** Feather quill crease · Radon transform · Decision fusion · Invariant moment · SVD

### 14.1 Introduction

Badminton is a labor-intensive products with about ten detection steps from feather selecting to finished badminton. In the whole steps, parameter extraction of feather quill (Referred to as “FQ”) is a key link for feather grading. Traditional detection methods exist disadvantage as follows: manual operation, high labor intensity, instability in sorting quality. At present, we have do some study of the problem [1–3]. Due to slender structure of FQ with variable width, camber and

---

H. Yue (✉) · R. Wang · J. Zhang · Z. He  
Faculty of Automation, Guangdong University of Technology, Guangzhou, China  
e-mail: yuehongwei420@163.com

H. Yue  
Zhongkai University of Agriculture and Engineering, No 24 DongSha Road Hai Zhu,  
Guangzhou City, China

curvature, in addition, boundary between crease of FQ and background is fuzzy, this all lead to misdetection of the large number of good FQ. As most of FQs without creases in actual production, detection method requires not only effective feature extraction, but also a high degree of non-crease detection accuracy. The machine vision in detect defects of electronics manufacturing, machinery manufacturing, textile, metallurgy, paper, packaging and agricultural industries has a wide range of applications, and its detection algorithm is highly targeted does not have the versatility [4–8]. Because of the fuzzy boundary of the crease, this is a better choice for crease extraction and recognition based on regional shapes. Moments are often used to represent image features, such as Hu-moment [9], Zernike-moment [10–12], etc. However, these invariant moments have huge computation and are susceptible to noise interference.

The statistical analysis result of FQ crease shows that most of the FQ crease are straight-line segment with regular width which is approximately perpendicular to the radial physiological textures. In view of this feature, this paper firstly uses local-angle Radon transform to extract moment invariants of gray and horizontal gradient dual-mode of target region, then uses Singular Value Decomposition (SVD) to get feature vector of the two kinds of modal to eliminate the influence of physiological textures; secondly, according to the feature vector, two recognition results of crease can be get; finally, based on above two recognition results, the final recognition result of the image is achieved by the fusion of recognition results of the dual-mode at the decision level. The results of recognition experiment show that this method, which not only has better noise immunity ability but also reduces the rate of misjudgment of non-crease effectively, has practical application value.

## 14.2 Radon Transform Descriptor

By definition the Radon transform of an image is determined by a set of projections of the image along lines taken at different angles. Let  $f(x, y)$  be an image. Its Radon transform is defined by [13]:

$$R(\theta, t) = R\{f(x, y)\} = \int_{R^2} f(x, y)\delta(x \cos \theta + y \sin \theta - t) dx dy \quad (14.1)$$

where  $\delta(t)$  is the Dirac delta-function ( $\delta(t) = 1$  if  $x = 0$  and 0 elsewhere),  $\theta \in [0, \pi]$  and. In other words,  $R(\theta, t)$  is the integral of  $f$  over the line  $L_{(\theta, t)}$  defined by  $x \cos \theta + y \sin \theta = t$ . Consequently, the Radon transform of an image is determined by a set of projections of the image along lines taken at different angles. An image recognition framework should allow explicit invariance under the operations of translation, rotation, scaling. But it will be difficult to recover all the parameters of the geometric transformations from the Radon transform [see Eq. (14.1)]. To overcome this problem, we propose an Radon transform [14].

Let the following transform be:

$$R_\alpha = R(\theta, t, \alpha) = \frac{\int_{-\infty}^{\infty} (R(\theta, t))^\alpha dt}{(A_f)^{\frac{\alpha+1}{2}}} \quad (14.2)$$

where  $A_f = \int_{-\infty}^{\infty} R(\theta, t) dt = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy$ . We can show the following properties.

Translation of a vector  $\vec{u} = (x_0, y_0) : g(x, y) = f(x + x_0, y + y_0)$

$R\{g(x, y)\} = R(\theta, t + x_0 \cos \theta + y_0 \sin \theta)$ . Substituting Eq. (14.2), we obtain:

$$\begin{aligned} R_g(\theta, t, \alpha) &= \frac{\int_{-\infty}^{\infty} (R_f(\theta, t + x_0 \cos \theta + y_0 \sin \theta))^\alpha dt}{(A_g)^{\frac{\alpha+1}{2}}} = \frac{\int_{-\infty}^{\infty} (R_f(\theta, t'))^\alpha dt'}{(A_f)^{\frac{\alpha+1}{2}}} \\ &= R_f(\theta, t, \alpha) \end{aligned}$$

Scaling of  $\lambda$ :

$g(x, y) = f(\frac{x}{\lambda}, \frac{y}{\lambda}) : A_g = \lambda^2 A_f, R\{g(x, y)\} = \lambda R_f(\theta, \frac{t}{\lambda})$ .: Substituting Eq. (14.2),

we obtain:

$$R_g(\theta, t, \alpha) = \frac{\int_{-\infty}^{\infty} (R_f(\theta, \frac{t}{\lambda}))^\alpha dt}{(A_g)^{\frac{\alpha+1}{2}}} = \frac{\int_{-\infty}^{\infty} (\lambda R_f(\theta, \frac{t}{\lambda}))^\alpha dt}{(\lambda^2 A_f)^{\frac{\alpha+1}{2}}} = R_f(\theta, t, \alpha),$$

Rotation by  $\theta_0 : R_f(\theta, t, \alpha) = R_f(\theta + \theta_0, t, \alpha)$ .

The area of image  $A_f$  can be calculated using any  $\theta$ . To summarize, the  $R_\alpha$  is invariant moment under translation and scaling if the transform is normalized by a scaling factor  $\alpha (\alpha \in \mathbb{Z}^+, \alpha > 1)$ . A rotation of the image by an angle  $\theta_0$  implies a shift of the Radon transform in the variable  $\theta_0$ . In the next section we propose an extension to solve this drawback.

### 14.3 Proposed Scheme

Taking on different values of scaling factor  $\alpha$  [see Eq. (14.2)], matrix invariants  $R = [R_2, R_3, \dots, R_\alpha]$  can be constructed,  $R_i (2 \leq i \leq \alpha)$  is  $i$  order invariant moment. In terms of matrix invariants, we particularly focus on Singular Value Decomposition (SVD) [15] to extract algebraic features which represent intrinsic attributions of an image. The SVD is defined as follows:  $R = U \sum V^T$ , where  $R$  is an  $m \times n$  real matrix,  $U$  is an  $m \times m$  real unitary matrix, and  $V$  is an  $n \times n$  real unitary matrix.  $\sum = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n, 0, \dots, 0)$  is an  $m \times n$  diagonal matrix containing singular values  $\lambda_1 \geq \lambda_2 \geq \dots, \geq \lambda_n > 0$ : let  $\eta = (\lambda_1, \lambda_2, \dots, \lambda_n)$  be invariant feature vector of an image.

Raw material for badminton usually is duck feather or goose feather with radial physiological textures which is approximately perpendicular to FQ crease. However, using SVD can eliminate the differences of rotated image and lead to

confusion physiological texture and creases. From Eq. (14.2), we can know Radon transform is determined by a set of projections at angle  $\theta \in [0, \pi]$ . The proposed scheme, called local-angle Radon transform (Referred to as “LR transform “), is LR transform is determined by a set of projections at orientation angle  $\theta \in [0^\circ, 15^\circ] \cup [165^\circ, 180^\circ]$ . This method by reducing the angle can eliminate the impact of physiological texture.

## 14.4 Crease Detection Algorithm

In order to improve the detection accuracy, sidelight image with side-lighting is operated. Below is a brief summary of this algorithm. Assume getting the target region as suspected crease sub-image  $f_i$  after pretreatment. Then calculating the LR transform of  $f_i$  in gray domain and horizontal gradient domain. Combining with SVD, we can get recognition results  $r_{gray}$  and  $r_{grads}$  of gray and grads dual-mode, respectively. Assume also a set of  $r_{gray} = \{0, 1\}$ ,  $r_{grads} = \{0, 1\}$ , where 0 represents crease, 1 represents non-crease. For decision level fusion is a high-level integration with better anti-interference ability and fault tolerance and can effectively reflect the different types of information for each side of the target.

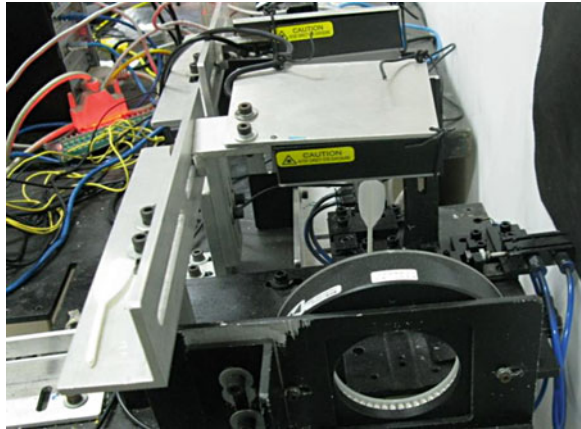
After getting two recognition results  $r_{gray}$  and  $r_{grads}$  of dual-mode, using decision level fusion method can obtain the final recognition result  $f_r \cdot f_r$  can be obtained by  $r_{gray}$  and  $r_{grads}$  as follows:  $f_r = \begin{cases} 1, & f_{gray}(i) + r_{grads}(i) = 2 \\ 0, & else \end{cases}$ , where 0 represents crease, 1 represents non-crease. The entire algorithm is defined as follows:

- (1) Compute LR transform of  $f_i$  in gray and horizontal gradient domain.
- (2) Construct matrix invariants  $R$ .
- (3) Using SVD to extract invariant features  $\eta_i$  and compute  $d_i = \|\eta_i\|$ .
- (4) Label  $r_{gray} = 0$  or  $r_{grads} = 0$ , if  $d_i > k$  and 1 otherwise;  $k$  is experience value.
- (5) Compute  $f_r$ .

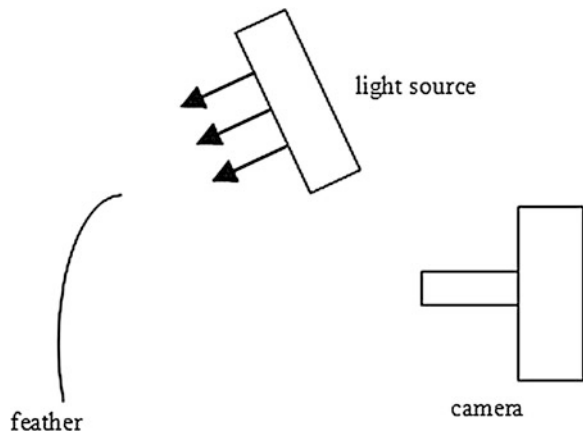
## 14.5 Analysis of Experimental Results

In this paper, experimental samples are collected from feather detection system, and raw material is duck feather. The detection system through CCD camera obtains feather image with side-lighting. Homemade feather acquisition system and its corresponding schematic diagram are shown in Figs. 14.1 and 14.2. From sidelight image (Fig. 14.3), we can get FQ image (Fig. 14.4) by segmentation technique. The testing database contains 356 non-crease sub-images and 79 crease sub-images. During the pretreatment of FQ, they have been obtained suspected

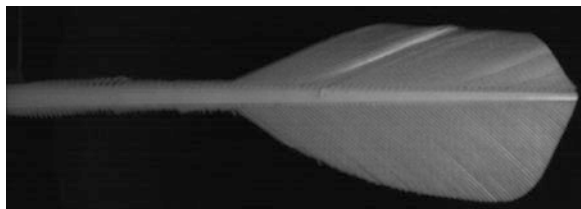
**Fig. 14.1** Acquisition system of feather



**Fig. 14.2** Side-lighting diagram



**Fig. 14.3** Sidelight image



**Fig. 14.4** Feather quill image





**Fig. 14.5** Experimental samples

crease sub-images, and have been confirmed and compared with physical objects. Because of slender structures and camber of FQ, these sub-images have abandoned edge in order to reduce effects of edge plus noise.

### 14.5.1 Comparison of Feature Vector

The experimental samples are shown in Fig. 14.5. Figure 14.5a is a non-crease sub-image; Fig. 14.5b is crease image of FQ root by noise interference, and the crease just only is half of the width; the crease's width of Fig. 14.5c is smaller; the crease of Fig. 14.5d has certain inclination and is not exactly perpendicular to the horizontal direction. After the pretreatment of the normalized Fig. 14.5, feature vector is obtained using Hu moment, Zernike moment, LR transform. The LR transform is performed in gray domain and gradient domain. Projection angle interval of LR transform is 1 degree and singular value as the recognition feature is obtained by 2, 3, 4 order invariant moments, as shown in Table 14.1. Because of limited space, the Table 14.1 gives only the first two singular values. As can be seen from the table, the distinction of singular value obtained by proposed method in gray and gradient domain is clear; and singular value obtained by Hu moment, Zernike moment have no obvious difference. Table 14.2 shows the singular value after the common Radon transform. Through contrasting singular value with Table 14.1, we can know singular value get larger because of the influence of radial physiological texture in the gray domain; and in horizontal gradient domain, the distinction between crease and no-crease gets smaller. The results show LR transform is more suitable for crease recognition than Radon transform.

**Table 14.1** Singular value of different methods

Test image		Hu moment	Zernike moment	LR transform	
				Gray	Gradient
Figure 14.5a	Value 1	0.1812	0.1770	0.011	0.1989
	Value 2	0.0047	0.0906	0.00001	0.0002
Figure 14.5b	Value 1	0.1815	0.2078	0.065	0.7104
	Value 2	0.0053	0.0558	0.0001	0.0088
Figure 14.5c	Value 1	0.217	0.1841	0.7988	1.5145
	Value 2	0.0181	0.1808	0.0016	0.0085
Figure 14.5d	Value 1	0.1886	0.2026	0.3087	0.9207
	Value 2	0.0079	0.0650	0.0010	0.0044

**Table 14.2** Results of radon transform

		Figure 14.5a	Figure 14.5b	Figure 14.5c	Figure 14.5d	
Radon transform	Gray domain	Value 1	3.0441	5.1224	2.5158	5.4878
		Value 2	0.0175	0.0448	0.1351	0.0171
	Gradient domain	Value 1	0.4815	0.6757	2.1708	0.8497
		Value 2	0.01	0.0209	0.0225	0.0101

**Table 14.3** Fusion result

	Crease/ %	Non-crease/ %
Gray-rate	86.41	90.12
Gradient-rate	85.29	88.24
After fusion	72.71	98.32

### 14.5.2 Recognition Comparison

In this paper, we use recognition rate index to measure performance for crease and non-crease recognition fusion before and after. Table 14.3 has presented crease and non-crease recognition rate by the LR transform. From the table, we can remark that the crease recognition rate has declined after fusion in gray domain (hereinafter referred to as gray-rate) and in gradient domain (hereinafter referred to as gradient-rate); non-crease recognition rate after fusion has big improvement over the previous two. Amount of non-crease feather is far more than amount of crease feather in the actual production, so the proposed fusion strategies to improve the non-crease recognition rate meets the demand for industrial production.

## 14.6 Conclusion

This paper puts forward a new method of FQ crease recognition. It uses local-angle Radon transform to extract invariant moment of FQ combining with SVD to get features (singular value) which have invariance in translation and scale to eliminate the disruption of radial physiological texture. Comparing to Hu moment and Zernike moment, the experimental results show that this method has good robustness and better distinguish effect. Finally decision fusion achieves high recognition rate of non-crease. The method has some value for on-site testing with 60 ms average running time in vc++ 6.0 environment.

## References

1. Liu H, Wang R, Li X (2011) Improved algorithm of feather image segmentation based on active contour model. *J Comput Appl* 32(8):2246–2245 (in Chinese)
2. Liu H, Wang R, He Z (2011) Slender object extraction for image segmentation based on centerline snake model. *Opto Electron Eng* 38(9):124–129 (in Chinese)



3. Liu H, Wang R, Li X (2011) The finite ridgelet transform for defect detection of quill. *ADME2011* 9:931–936 (in Chinese)
4. Ruan J (2009) Research on non-planar measurement of surface roughness based on texture index from grey level co-occurrence matrix. Yantai University, Yantai (in Chinese)
5. Zheng XL, Li M, Luo HY et al (2013) Application of a new method based on mathematical morphology in brain tissue segmentation. *Chinese J Sci Instrum* 31(2): 464–469 (in Chinese)
6. Subirats P, Dumoulin J, Legeay V et al (2006) Automation of pavement surface crack detection using the continuous wavelet transform. In: 2006 international conference on image processing, Florence, Italy, pp 3951–3954
7. Zhang H, Wu Y, Kuang Z (2009) An efficient scratches detection and inpainting algorithm for old film restoration. In: *Proceedings of the 2009 international conference on information technology and computer science*. IEEE Press, Washington, DC, pp 75–78
8. Sorncharean S, Phiphobmongkol S (2008) Crack detection on asphalt surface image using enhanced grid cell analysis. In: 4th IEEE international symposium on electronic design, test and applications, Hong Kong, pp 49–54
9. Hu MK (1962) Visual pattern recognition by moment invariants. *IEEE Trans Inf Theor* 8(1):179–187
10. Teh CH, Chin RT (1988) On image analysis by the methods of moments. *IEEE Trans Pattern Anal Mach Intell* 10(4):496–512
11. Kan C, Srinath MD (2002) Invariant character recognition with Zernike and orthogonal fourier-mellin moments. *Pattern Recogn* 35(1):143–154
12. Liao SX, Pawlak M (1998) On the accuracy of Zernike moments for image analysis. *IEEE Trans Pattern Anal Mach Intell* 20(12):1358–1364
13. Deans SR (1983) *The radon transform and some of its applications*. Wiley, New York
14. Lv Y (2008) *Shape affine invariant feature extraction and recognition*. National Defense Science and Technology University, Changsha (in Chinese)
15. AlShaykh OK, Doherty JF (1996) Invariant image analysis based on radon transform and SVD. *IEEE Trans Circ Syst II Analog Digital Signal Process* 43(2):123–133

# Chapter 15

## A New Vision Inspired Clustering Approach

Dequan Jin and Zhili Huang

**Abstract** In this paper, a new clustering approach by simulating human vision process is presented. Human is good at detecting and segmenting objects from the background, even when these objects have not been seen before, which are clustering activities in fact. Since human vision shows good potential in clustering, it inspires us that reproducing the mechanism of human vision may be a good way of data clustering. Following this idea, we present a new clustering approach by reproducing the three functional levels of human vision. Numeric examples show that our approach is feasible, computationally stable, suitable to discover arbitrarily shaped clusters, and insensitive to noises.

**Keywords** Neural field theory · Clustering Analysis · Amari's model · Stationary solution · Dynamical system

### 15.1 Introduction

Clustering is primitive in human learning activities [1]. All the unsupervised learning activities can be considered as clustering activities [2]. The most fundamental human clustering activity may be in human vision, for instant, in object

---

D. Jin (✉)  
School of Mathematics and Information Science, Guangxi University,  
No. 100 Daxue Road, Nanning, Guangxi, China  
e-mail: dqjin@yahoo.cn

D. Jin  
School of Mathematics and Statistics, Xi'an Jiaotong University,  
No. 28 West Xianning Road, Xi'an, Shaanxi, China

Z. Huang  
School of Mechanical Engineering, Guangxi University,  
No. 100 Daxue Road, Nanning, Guangxi, China

detection or segmentation, which are elementary visual activities. Human vision is highly complex. A simple visual activity may involve a lot of neuronal structures. If we just investigate one or several ones of these structures, it may be not enough to illustrate the visual activity. But if we consider the whole visual system that includes most of these coupled structures, it may become too complicated to be analyzed. Instead of considering individual neuronal structure, some investigations on visual perception indicates a feasible way to avoid these difficulties, which reproduce the functional levels of human vision, each of which might be structurally distributed [3, 4].

In this paper, a new clustering approach is presented, by considering data as an image and clustering them by three-level neural field system for visual perception. Method for determining the range of excited regions in the activation distribution of neural field is also introduced.

## 15.2 Levels for Human Vision

Generally speaking, the whole procedure of human vision contains three functional levels at least, including the transfer level, the planning level, and the motor control level.

### 15.2.1 Transfer Level

In this level, eyes and some low-level neuronal structures are involved. Eyes act as sensors, whose main task is to accept the light intensity distribution and transform them into neural signals distributions. The visual information may be subjected to some spatial and temporal transformations in this level induced by retina and some neuronal structures [4].

Suppose the objects are static in visual field. Then the light intensity distribution is usually presented as a static image that consists of  $N$  light points  $\{x_i\}_{i=1}^N$ , which can be described as [5]

$$I(z) = \frac{1}{N} \sum_{i=1}^N \delta(z - x_i) \quad (15.1)$$

where

$$\delta(z - x_i) = \lim_{\sigma \rightarrow 0} g(z, \sigma) \quad (15.2)$$

$g(z, \sigma)$  is a Gaussian function

$$g(z, \sigma) = \exp\left(-\frac{\|z\|^2}{\sigma^2}\right) \quad (15.3)$$

The visual information  $I(z)$  may be subject to some filtering effects in the transfer level, the most significant one of which in object detection and segmentation is the blurring effect induced by the retina, which is usually described as the Gaussian filtering process. Then the output neural signal distribution  $S(z)$  is given by

$$S(z) = S(z, \sigma_{in}) = \int_{\Omega} I(z - z')g(z, \sigma_{in})dz' \quad (15.4)$$

$\sigma_{in}$  is a scale parameter that can be understood as the distance between object and eye, or the curvature of crystalline lens [6]. If light points are too close that under the resolution of eyes, they cannot be identified in  $S(z)$  individually.  $S(z)$  is the output of the transfer level, as well as the input of the planning level.

### 15.2.2 Planning Level

In this level, the input neural signal  $S(z)$  would be processed by neurons. A visual perception is presented in the form of activation distribution of these neurons.

There are many neural models in describing the activity of neurons. A popular one of them is the Amari's dynamical neural field [7]:

$$\tau \dot{u}(z, t) = -u(z, t) + \int_{\Omega} w(z, z')\theta(u(z', t))dz' + S(z) - h \quad (15.5)$$

The vector space  $\Omega$  is called perceive space.  $\tau$  is a positive time constant.  $h$  is the resting level parameter. The region  $\{z \in \Omega : u(z, t) > 0\}$  is called excited region, denoting the activated neurons. The excited region usually corresponds to a perceived pattern.  $\theta(u)$  is a monotonically increasing nonlinear threshold function satisfying that  $\lim_{u \rightarrow -\infty} \theta(u) = 0$  and  $\lim_{u \rightarrow +\infty} \theta(u) = 1$ , for instant, the step function. It describes the neural field feedback of each excited point to its neighboring positions in  $\Omega$  with an interaction strength which is determined by interaction function  $w(z, z')$ .

Mostly, the interaction function  $w(z, z')$  is isotropic and usually written as  $w(z - z')$ . In this case,  $w(z)$  is also called as the interaction kernel of the neural field. Approximating the neurophysiologic lateral interaction among neurons, the lateral interaction of neural field is usually assumed to be locally exciting and globally inhibiting. One of the typical interaction kernels is the difference of Gaussian (DoG) functions with constant inhibition, given by

$$w(z) = Ag(z, \sigma) - Bg(z, \gamma\sigma) - h_{ker} \quad (15.6)$$

where  $\gamma > 1$ .

There are three important types of stable solutions to system (15.5), which are  $\phi$ -solution, “bubble”-solution and  $\infty$ -solution:

1. An equilibrium solution  $u^*(z)$  is called  $\phi$ -solution if  $u^*(z) \leq 0$  for all  $z \in \Omega$ ;
2. An equilibrium solution  $u^*(z)$  is called “bubble”-solution if  $u^*(z) > 0$  for  $z$  in a subset  $D \subset \Omega$ ;
3. An equilibrium solution  $u^*(z)$  is called  $\infty$ -solution if  $u^*(z) > 0$  for all  $z \in \Omega$ .

### 15.2.3 Motor Control Level

Motor control level aims at sending out control signals to specific organs, for instant, eyes, according to the neuron activity  $u(z, t)$  which is the output the planning level. For different purposes, this level would have different descriptions. For instant, in the investigation on saccadic motor planning, to control eyes to stare at an object in visual field, let  $z^* = \frac{P}{M}$ , where  $P = \int_{\mathbb{R}} z\theta(u^*(z))dz$  and  $M = \int_{\mathbb{R}} \theta(u^*(z))dz$ , then  $z^*$  is the density center of activation distribution which corresponds to the center of object [4].

## 15.3 Clustering Approach Based on Vision

Since human vision shows good potential in clustering, it is possible to find a feasible clustering approach by simulating visual mechanisms. In this section, we present a new clustering approach by reproducing the three levels of vision. Some numeric examples are given to show the feasibility and advantages of our approach.

### 15.3.1 Transfer Level for Clustering

In this level, the first thing is to transform data set  $X = \{x_i \in \mathbb{R}^n : i = 1, 2, \dots, N\}$ , to image which can be accepted by visual system by

$$I(z) = \frac{1}{N} \sum_{i=1}^N \delta(z - x_i) \quad (15.7)$$

In this way, we obtain a data distribution  $I(z)$ .

The data distribution  $I(z)$  is transformed into neural input distribution  $S(z)$  by a Gaussian filtering process:

$$S(z) = S(z, \sigma_{in}) = \int_{\Omega} I(z - z')g(z, \sigma_{in})dz' \quad (15.8)$$

### 15.3.2 Planning Level for Clustering

The aim of this level is to discover clusters in data, i.e., produce a perception of clusters, according to the neural signal distribution  $S(z)$ .

Since Amari's model achieves successes in illustrating phenomena in visual perception, we also employ it in the planning level for clustering, which is given by (15.5):

$$\tau \dot{u}(z, t) = -u(z, t) + \int_{\Omega} w(z, z')\theta(u(z', t))dz' + S(z) - h \quad (15.9)$$

where  $\Omega \subset \mathbb{R}^n$ .

As soon as  $S(z)$  being transferred to the planning level, the neural field begins to evolve, until the field reaches its steady state  $u^*(z)$ . Several bubbles, i.e., excited regions may be sustained in  $u^*(z)$ , whose number and range generally depends on the input  $S(z)$ , the kernel  $w(z)$  and the resting level  $h$  of the neural field. By grouping the data located in the same connected excited region into a cluster, the clusters of the data set  $X$  are perceived.

### 15.3.3 Motor Control Level for Clustering

In visual perception, the motor control level sends out control signal based on the activation distribution  $u^*(z)$  given by the planning level, so that people can react based on their perception in the planning level, corresponding to outside visual stimulus. In the motor control level, we introduce some methods to point out the range of connected excited regions in  $u^*(z)$ .

When a connected excited region in  $u^*(z)$  is convex, its range is equal to the attraction domain of a corresponding equilibrium point of the gradient dynamical system

$$\frac{dz}{dt} = \nabla u^*(z) \quad (15.10)$$

By estimating its corresponding attraction domain, we can estimate the range of a connected excited region. A feasible way for estimating the domain of attraction of such a system is presented in [8], which employs an iterative expansion approach. Details can be seen in [8].

## 15.4 Algorithm and Examples

On the basis of the above strategies, for a data set  $X = \{x_i \in \mathbb{R}^n : i = 1, 2, \dots, N\}$ , we present a clustering approach as following:

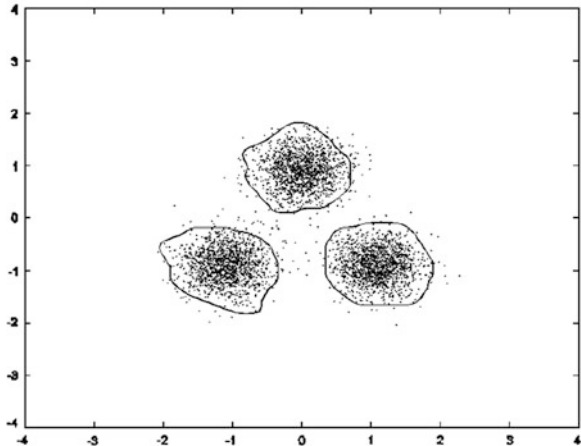
1. Select a scale  $\sigma_{in} > 0$ , and the interaction kernel  $w(z)$ . Compute the signal distribution  $S(z)$  by (15.8);
2. Let  $u(z, 0) = -h$  for  $z \in \Omega$ . Compute the steady state  $u^*(z)$  of system (15.9). If there are  $m$  excited regions, take all the data points in the same connected excited region into a cluster, denoting by  $C_j$ ,  $j = 1, 2, \dots, m$ . If there are unlabeled data points, go to step 3; else, let  $M = m$ , go to step 4.
3. If the unlabeled data points locate in excited regions, then group them to the clusters corresponding to these excited region; else, if the unlabeled data points locate in some peaks with negative activation, group the data locating in the same peaks into new clusters  $C_{m+i}$ ,  $i = 1, 2, \dots, \tilde{m}$ . Let  $M = m + \tilde{m}$ .
4. Let  $Sc = \{C_j\}_{j=1}^M$ , then  $Sc$  is the clustering result.

To show the feasibility of our approach, we give some numeric examples as shown in Figs. 15.1 and 15.2. The kernel  $w(z)$  is given by (15.6). Parameters are given as  $\sigma = 0.07$ ,  $\gamma = 1.1$ ,  $\tau = 0.1$ ,  $A = 1.2$  and  $B = 0.1$ ,  $h_{ker} = 0.002$ ,  $h = 0.02$ .

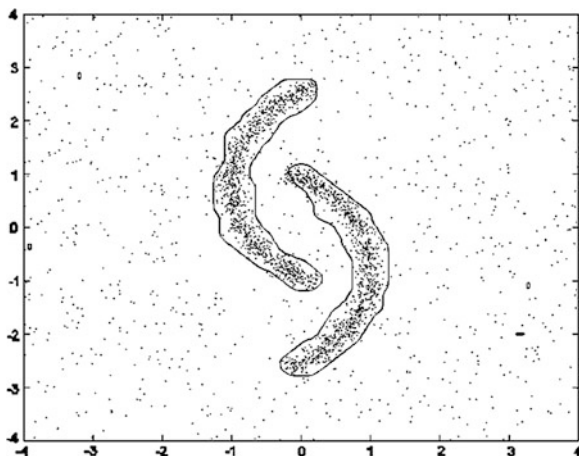
As shown in these examples, it can be seen that our approach has some advantages, for instant, our approach doesn't require the number of clusters and specific learning step, is suitable to discover clusters with arbitrary shape. Noises and isolated points are easy to be identified in the above results.

Since convolution is involved in Amari's neural field, which consumes a lot of computation, the computational time expense of this approach is high. As a result, limited by current computer technology, this approach cannot deal with high dimensional clustering problems efficiently. However, these Numeric examples show that this approach has high accuracy in clustering and anti-noise ability.

**Fig. 15.1** Three clusters obtained by this approach for three-Gaussian data set



**Fig. 15.2** Two non-convex clusters obtained by this approach for “Double C” data set with noises



Moreover, these clustering results are highly close to human cognition. So this approach shows potential especially when breakthroughs in computer technology like quantum computer are made in future.

## 15.5 Conclusion

In this paper, we present a new clustering approach inspired from human vision. By reproducing the mechanism of the three functional levels of human vision, we present a new clustering approach. This approach is biologically plausible, robust to noises and suitable to discover arbitrary shaped clusters. To show the feasibility of our approach, some numeric examples are given.

Nevertheless, our approach is an attempt. Our approach is suitable for all kinds of data sets theoretically, but it relies on a neural field which contains a convolution on its right hand side, which would consume much computer time.

**Acknowledgments** This paper is sponsored by the Scientific Research Foundation of Guangxi University (Grant No. XBZ120366) and supported by NSFC, Tian Yuan Special Foundation, Project No. 11226141.

## References

1. Xu R, Wunsch D (2005) Survey of clustering algorithms. *IEEE Trans Neural Netw* 16(3):645–678
2. Bishop CM (2006) *Pattern recognition and machine learning*. Springer, New York
3. Giese MA (1999) *Dynamic neural field theory for motion perception*. Kluwer Academic Publishers, Norwell



4. Kopecz K, Schoner G (1995) Saccadic motor planning by integrating visual information and pre-information on neural dynamic fields. *Biol Cybern* 73(1):49–60
5. Leung Y, Zhang J, Xu Z (2000) Clustering by scale-space filtering. *IEEE Trans Pattern Anal Mach Intell* 22(12):1396–1410
6. Xu Z, Meng D, Jing W (2005) A new approach for classification: visual simulation point of view. *Lecture notes in computer science advances in neural networks—proceedings of international symposium on neural networks, ISNN 2005, Part II, Chongqing*, pp 1–7
7. Amari S (1977) Dynamics of pattern formation in lateral-inhibition type neural field. *Biol Cybern* 27(2):77–87
8. Jin D, Peng J (2009) A new approach for estimating the attraction domain for Hopfield-type neural networks. *Neural Comput* 21(1):101–120

# Chapter 16

## Multiclass Vehicle Tracking Based on Local Feature

Zhiming Qian, Jiakuan Yang and Lianxin Duan

**Abstract** This paper presents a real time multiclass vehicle tracking method. The method uses a combination of machine learning and feature analysis to track the vehicles on the road. Multiclass SVM are trained using train samples to achieve detection and classification of vehicles in video sequences of traffic scenes. The detection results provide the system used for tracking. Each class vehicle is tracked by SIFT method. Experimental results from highway scenes are provided which demonstrate the effectiveness of the method.

**Keywords** Vehicle detection · Vehicle tracking · Feature analysis

### 16.1 Introduction

Video based intelligent transportation systems (ITS) are getting large attention as an attractive field, not only because they are easy to install and operate, but also because they have the potential to provide a much richer description about vehicle. As the basic parts, tracking of vehicle is a fundamental problem in ITS. For this task, we need to first detect the vehicle and segment them from the video images, and then track them across different frames while maintaining the correct identities.

Robust tracking of vehicles on the road based on video is a challenging problem. Roads are dynamic environments, with the illumination and background changes. The sizes and the locations of vehicles on the road are diverse. There is high variability in the appearance of vehicles with viewpoint, illumination, and

---

Z. Qian (✉) · J. Yang · L. Duan  
Chuxiong Normal University, Chuxiong 675000, China  
e-mail: qzhiming@126.com

possible articulation. Moreover, partial occlusion of vehicles of interest by other vehicles or objects on the road is also an important factor influencing tracking.

For the last two decades researchers have spend quality time to develop different methods that can be applied in the field of video based vehicle tracking [1–6]. The existing tracking approaches may be classified into four major categories:

1. Region based method: this method subtracts image frame containing vehicles from the background frame which is then further processed to obtain vehicle regions (blobs). Then these vehicle regions are tracked. It can work well in free flowing traffic conditions, but the disadvantage is that it has difficulty in handling shadows and occlusion.

2. Active contour based method: this method represents vehicle by bounding contour of the object and dynamically update it during the tracking. The advantage of active contour tracking over region-based tracking is the reduced computational complexity. But the disadvantage of the method is their inability to accurately track the occluded vehicles and tracking need to be initialized on each vehicle separately to handle occlusion better.

3. Feature based method: this method extracts suitable features from the vehicle regions and these features are processed to track the vehicles correctly. The method has low complexity and also can handle occlusions well. The disadvantage is the recognition rate of vehicles using two-dimensional image features is low, and the problem that which set of sub features belong to one object is complex.

4. Model based method: this method tracks vehicle by matching a projected model to the image data. The advantages of model based vehicle tracking is it is robust to interference between nearby images and also be applied to vehicle classification. But the method has high computational cost and they need detailed geometric object model to achieve high tracking accuracy.

In real world, the vehicle type is various. Comparing the strategy that all vehicles are categorized as single class, multiclass vehicle tracking have great practical significance and applicable value great practical importance. In this paper, a method for video multiclass vehicles tracking is introduced. The proposed method has the following characteristics: (1) It has multiclass vehicles detection ability; (2) It can track multiclass vehicles accurately in real-time environment. The proposed method in this paper has been validated with video vehicle sequences from real-world traffic scenes.

## **16.2 The Proposed Method**

### ***16.2.1 Overall Structure***

Given an input of a video sequence taken from roadway vehicles, system first outputs the types and locations of the vehicles in the images, then a feature information description of the detected vehicles is obtained, and finally this

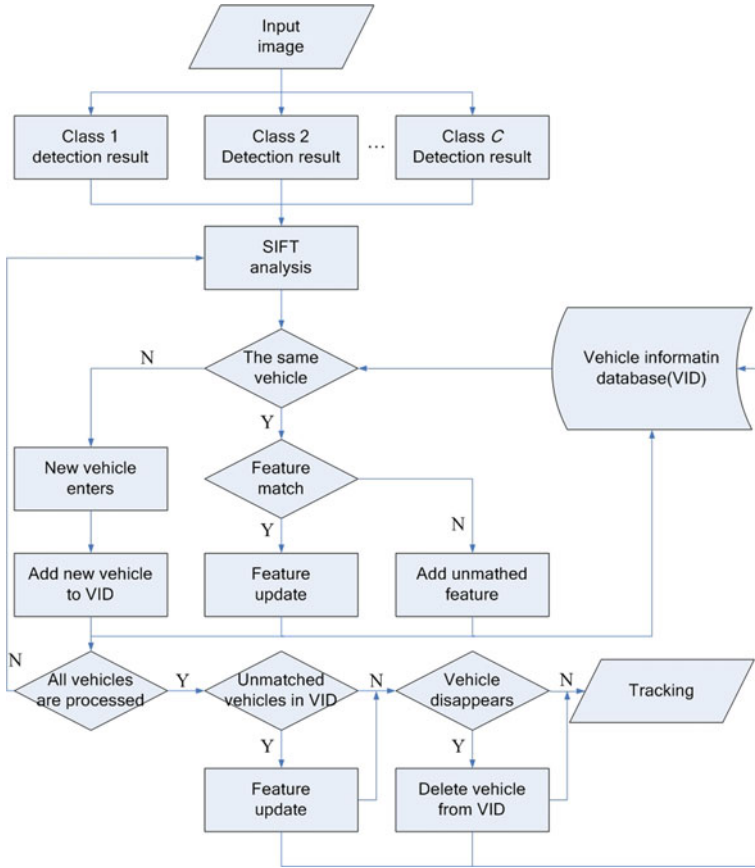


Fig. 16.1 Flowchart of multiclass vehicle tracking

description is used to match the detected vehicles in the next frame. The framework contains three main processes: vehicle detection, vehicle classification, and vehicle tracking. A general overview of the system flowchart can be seen in Fig. 16.1.

### 16.2.2 Vehicle Detection and Classification

In order to achieve vehicle detection and classification task, multiclass SVM is employed to our framework. The SVM has been introduced as one of the most efficient learning algorithms in computer vision. To be useful, the task of vehicle classification should categorize vehicles into a sufficiently large number of classes. However as the number of class increases, the processing time required also

increases. Therefore, a simple classification method is needed which can quickly categorize vehicles at a coarse level. Based on the application, further classification can be done. In the paper, we use the “one against one” method in the LIBSVM [7] to learn Haar wavelet features for vehicle classification.

The one-against-one method constructs an SVM for every pair of classes by training it to discriminate the two classes. If  $k$  is the number of classes, then  $k(k-1)/2$  classifiers are constructed and each one trains data from two classes. The decision function for class pair  $ij$  is defined by

$$f_{ij}(x) = (\phi(x) \cdot w^{ij}) + b^{ij} \quad (16.1)$$

It is found by solving the following optimization problem:

$$\min \frac{1}{2} \|w^{ij}\|^2 + C \sum_n \xi_n^{ij} \quad (16.2)$$

$$\begin{cases} \phi(x_n) \cdot w^{ij} + b^{ij} \geq 1 - \xi_t^{ij}; & \xi_t^{ij} \geq 0, \quad x_t \text{ in the } i\text{th class} \\ \phi(x_n) \cdot w^{ij} + b^{ij} \leq \xi_t^{ij} - 1; & \xi_t^{ij} \geq 0, \quad x_t \text{ in the } j\text{th class} \end{cases} \quad (16.3)$$

Finally, the “max wins” voting strategy is used to determine the class of a test pattern in this approach.

### 16.2.3 SIFT Feature Analysis

SIFT (Scale Invariant Feature Transform) is a well-established local feature descriptors, which was proposed in 1999 by Lowe [8]. Duo to SIFT feature descriptor is invariant to uniform scaling, orientation, and partially invariant to affine distortion and illumination changes, it has been widely applied to object tracking and image matching. For multiclass vehicle tracking, we need a kind of feature which can describe different vehicles accurately, the SIFT feature is very suitable in the circumstance. Main process of the SIFT algorithm is as follows:

Interest points for SIFT features correspond to local extrema of difference-of-Gaussian filters at different scales. Given a Gaussian-blurred image described as the formula

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (16.4)$$

where  $L$  is the scale space of an 2D image,  $I(x, y)$  is the gray value of input image in the coordinates  $(x, y)$ ,  $G(x, y, \sigma)$  is a variable scale Gaussian, whose result of convolving an image with a difference-of-Gaussian filter is given by

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (16.5)$$

which is just be different from the Gaussian-blurred images at scales  $\sigma$  and  $k\sigma$ . Interest points are identified as local maxima or minima of the DoG images across

scales. Each pixel in the DoG images is compared to its 8 neighbors at the same scale, plus the 9 corresponding neighbors at neighboring scales. If the pixel is a local maximum or minimum, it is selected as a candidate feature point. Remove the low contrast candidate points and eliminated the edge response, then use Hessian matrix to compute the principal curvatures and eliminate these feature points that have a ratio between the principal curvatures greater than the ratio.

Finally, an orientation histogram was formed from the gradient orientations of sample points within a  $4 \times 4$  region with 8 orientations around the feature point in order to get an orientation assignment. So the descriptor of SIFT that was used is  $4 \times 4 \times 8 = 128$  dimensions.

### 16.2.4 Feature Matching and Updating

For each vehicle detected from multiclass detection framework, extract SIFT feature and establish vehicle information database (VID). The VID consists of four parts: vehicle class, vehicle number, vehicle location (rectangle coordinates) and SIFT feature point descriptor (feature priority, feature point coordinate, orientation and scale), each vehicle detected from multiclass detection framework is tracked in a new video frame sequences by separately comparing its feature point with the same class vehicle from the VID. The Euclidean distance is introduced as a similarity measurement of feature characters.

Suppose  $N_i$  as the feature number of the current vehicle matching the  $i$ th vehicle of the VID,  $N$  as the total feature number of the current vehicle, the matching rate between the current vehicle and the  $i$ th vehicle of the VID can be defined as  $P_i = N_i/N$ . Set the threshold  $T$  for the matching parameters. When  $P_i$  is greater than  $T$ , the current vehicle is considered equivalent to matching the  $i$ th vehicle. Supposing that  $M_j$  is the number of the  $j$ th class vehicle of the VID,  $\{P_{ij} (j = 1, \dots, M_j)\}$  is the matching results of the current vehicle and all vehicles of the VID with the same class, and  $n$  is the number of elements in the set  $\{P_{ij} | P_{ij} > T, j = 1, \dots, M_j\}$ . When  $n = 1$ , the  $i$ th vehicle is matching with the  $j$ th vehicle of the VID with the same class; when  $n > 1$ , we select  $\max(p_i)$  as the matching result.

The VID stores the data of vehicle which appears in the recent video sequences. It needs to be updated after one frame, input the current vehicle data and delete the data of the long term unmatched vehicle. We set a feature priority for each feature point of the VID in the vehicle information update process.

Suppose  $R_{ij}$  as the feature priority of the  $j$ th feature point of the  $i$ th vehicle, the specific update process is as following:

(1) Add new vehicle: if the current vehicle is not matching all the vehicle of VID with the same class, this vehicle will be considered as a new vehicle, add its information into the VID, and set its feature priority of all feature points  $R = R_{\max}$ .

(2) Update feature priority: if the current vehicle matches the  $i$ th vehicle of the VID, the information of the  $i$ th vehicle will be update, set the feature priority of

these matching feature points  $R_{ij} = R_{\max}$ , and use new coordinate of these matching feature points to replace original coordinate. In addition, the feature priority of unmatched feature points between the current vehicle and the  $i$ th vehicle is replaced with  $R_{ij} = R_{ij} - 1$ , the new feature points of unmatched feature are added into the VID. After all matches of the current frame are finished, if there are no matching vehicles to be found from the VID, all the feature priority of these vehicles will be replaced with  $R_{ij} = R_{ij} - 1$ . When a frame image is completely processed, the feature point whose feature priority is equal to zero will be removed from the VID.

(3) Delete vehicle: When a frame image is completely processed, the vehicle whose feature priority of feature point meets the following condition will be deleted from the VID.

$$R_1 + R_2 + \cdots + R_{All} < \theta \quad (16.6)$$

### 16.3 Experiment

We consider the samples from a profile viewpoint for vehicles, and all video sequences which are achieved a frame rate of about 20 fps were generated by shooting around Chuxiong city under highways conditions. All our experiments shown below on a standard PC (Intel Core2 Duo E7500 2.93 GHz with 2 GB RAM). Set the threshold  $\theta = 0.2$ ,  $R_{\max} = 5$ , the number of class  $C = 4$  (motorcycle, bus, truck, and car).

In the training phrase, the data set is the image segmentation data, where each class is a vehicle type collected from a  $32 \times 16$  region of a vehicle image. The training set consists of 1000 samples per class. Some training images are shown in Fig. 16.2. In the tracking test, if the classifiers obtain detection result which gives the desired location and identifier, the result will be considered as the correct tracking in current frame, otherwise the result will be considered as the incorrect tracking in current frame. Since there are no suitable methods to compare the



**Fig. 16.2** A subset of the training samples for the four classes

**Table 16.1** Tracking results on video sequences

Class	Tracked number	Vehicles not tracked	Average number of frames during tracking
Motorcycle	71	6	36
Bus	85	3	46
Truck	78	2	42
Car	82	4	31

**Fig. 16.3** Multiclass tracking results in the experimental sequences

multiclass tracking effect, we just test our method on test data. Table 16.1 shows the tracking results for our method. Some tracking results in the video sequences are shown in Fig. 16.3.

## 16.4 Conclusion

We have proposed a real-time vision framework that tracks multiclass vehicles in video sequences. The method by analyzing the SIFT feature of detected vehicles to achieve vehicle tracking. The method is able to run in real time with simple, low-cost hardware. Our experimental results demonstrate effective, multiclass vehicle tracking in real traffic environments by applying the proposed method. If new classes of vehicles or unfamiliar environments are encountered, the proposed method can adapt itself to the changes and track vehicles successfully.

**Acknowledgments** This work was supported by the Natural Science Foundation of Yunnan Province, China (No. 2011FZ187).

## References

1. Sun ZH, Bebi G, Miller R (2006) On-road vehicle detection: a review. *IEEE Trans Pattern Anal Mach Intell* 28(5):694–711
2. Trivedi MM, Gandhi T, McCall J (2007) Looking-in and looking-out of a vehicle: computer-vision-based enhanced vehicle safety. *IEEE Trans Intell Transp Syst* 8(1):108–120



3. Morris BT, Trivedi MM (2008) Learning, modeling, and classification of vehicle track patterns from live video. *IEEE Trans Intell Transp Syst* 9(3):425–437
4. Withopf D, Jahne B (2006) Learning algorithm for real-time vehicle tracking. In: *IEEE conference on intelligent transportation systems*
5. Huang L, Barth M (2010) Real-time multi-vehicle tracking based on feature detection and color probability model. In: *IEEE conference on intelligent vehicles symposium*
6. Sivaraman S, Trivedi MM (2010) A general active-learning framework for on-road vehicle recognition and tracking. *IEEE Trans Intell Transp Syst* 11(2):267–276
7. Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2(3):1–39
8. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vision* 60(2):91–110

# Chapter 17

## Simulating Virtual Plants Based on Genetic Algorithm and L-Systems

Weilong Ding, Chen Hu and Yuanwei Zhu

**Abstract** As an effective technique to dynamically simulate the morphological development of plants, L-system is widely used in the field of plant modeling and visualization. It is a key step to obtain the axiom, the productions, and their parameters before simulating a special target. Usually, it is difficult to get those parameters and rules. In order to avoid the blindness and low efficiency in the modeling process, we proposed an automatic algorithm to simulate the images of virtual plant based on Genetic Algorithm and L-systems in this article. The relative techniques of our methods, such as the encoded mode of genetic individuals, the design of the primary population, the design strategies of genetic operations, and the design of evaluation function of fitness are introduced in details. The results demonstrate the algorithm's capability of modeling various plants and ease of use by novice users.

**Keywords** L-system · Genetic algorithm · Plant morphology · Simulation

### 17.1 Introduction

Virtual plant [1, 2] now has become a hotspot in the field of computer graphics, agronomy, etc., and attracted many computer scientists, botanists and mathematicians devoting themselves to this field. The research scope of virtual plant may include: individual plants modeling [3], functional-structural plant modeling [4–7], forest modeling [8] etc. Using an appropriate morphological development model to generate virtual plants is a key step for the process-based model. L-system is a widely used model among the many existing morphological development models [9]. However, when using the L-system to model virtual plant, we need to pre-determine

---

W. Ding (✉) · C. Hu · Y. Zhu  
College of Computer Science and Technology, Zhejiang University of Technology,  
Hangzhou, China  
e-mail: wlding@zjut.edu.cn

the rule parameters required by the model, such as the axiom and production rules. Usually it is quite difficult and quite time-consuming to obtain these rules and parameters for modeling the specified plant shape. For novice users, they may be blind and inefficient to obtain the rule parameters of L-system with a manual way. In order to improve the efficiency of the plant simulation, it is essential to study how to obtain the parameters and rules for axiom and productions.

In recent years, genetic algorithms and gene expression programming technology have been applied to solve a series of complex problems, and achieved good results, such as combinatorial optimization, production scheduling, automatic control, and artificial life. Since the 1990s, many scholars try to use genetic algorithm and gene programming to extract the rules and parameters of the L-system model, in order to improve the relevance and efficiency of virtual simulation. For example, Jacob [10], Ochoa [11], Mock [12], Hornby and Pollack [13] study how to evolving parameter-less, bracketed DOL-systems, as these are suited to the most general evolutionary methods [14]. However, they do not involve in context sensitive, stochastic and timed L-systems. Kokai [15], Traler [16], Bian et al. [17], and Venter and Hardy [18] limit themselves to evolving parametric L-system to simulate different plant shapes. Moreover, to keep the bracket balance of bracketed L-system, the operators of mutation, selection, and crossover have many limitations, thus the variety of individuals are limited. For example, the operation of mutation can not be applied to the symbols '[' and ']'. The crossover points can only in the scope between '[' and ']'. The evaluation function of fitness for some works are simplistic or limited to highly specific domains [8]. And some researches are only to obtain several of types and shapes of plant images, so they don't have the fitness function in the realization process of the genetic algorithms. Thus the superior individuals which form the new population have to be chosen manually in each evolution process according to aesthetic standard for a next iteration. In terms of search efficiency, however, there are great limitations existed in such selection methods which are based on visual impression [14].

Although considerable works have been conducted in the field of virtual plant modeling based on Genetic Algorithm, many of them are still at the primary stage, and need to be improved with respect to various aspects, as we mentioned above. So in this study, an automatic algorithm to simulate the images of virtual plant based on Genetic Algorithm and L-systems is proposed, so as to avoid the blindness and low efficiency in the modeling process.

## 17.2 Related Techniques

### 17.2.1 L-Systems

L-system is a string rewriting system. It can be used to describe the occurrence and growth processes of plant morphology. The essence of it is a string rewriting system, which uses an axiom and a set of productions to generate a string by finite

recursive operations, and the string can be interpreted as the structure of a plant. L-system is suitable for describing morphologic architectures of plants. By the persistent efforts of many scholars, L-system has been successfully applied to a variety of occasions [9].

### ***17.2.2 Genetic Algorithm***

Genetic Algorithm (GA) is a search heuristic that simulating natural evolution process, which can be used to generate useful solutions on optimization and searching problems. It generates solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover. The core content includes individual coding, the initial population setting, the fitness function design, genetic operation design and parameter setting, etc. As one of the effective methods in optimization theory, genetic algorithm has applied in a series of complex optimization problems and achieved good results.

## **17.3 The Process to Modeling Specific Plants Based on GA**

### ***17.3.1 Coding Scheme and Population Initialization***

The symbol encoded is one of the more commonly used encoding. As we mentioned above, L-system is a string rewriting system, and its axiom and productions are composed by a number of characters. According to the characteristics of the characters in L-system production rules, and taking into account the efficiency and convenience of the crossover and mutation operations, we design the Gemini body coding scheme. A chromosome consists of two sub-chromosomes. One sub-chromosome is used to encode a string without brackets, and another chromosome is used to encode the brackets string. The start number of the gene is from 0. In a chromosome, if a gene whose value equals  $-1$ , it will not be displayed in the phenotype, and we call it “recessive gene”. In the process of encoding, the value of the recessive gene is set to  $-1$ , and the dominant gene is set to a non-negative integer. The individual’s phenotype is determined by a non-negative dominant gene. When decoding the a chromosome into a production string, we just need to fill the gene string after removed “recessive gene” into the bracket string successively. And then according to the defined relationship between the integers and the characters, a chromosome is mapped to a sequence of characters.

### ***17.3.2 Initial Population***

The quality of the initial population is one of the key factors affecting the results of genetic algorithm. The generation of individual consists of two parts: L-system productions and the number of iterations. Taking into account the two aspects of the individuals generated with high fitness and the diversity of the gene, the initial population is divided into two parts: based on preset individuals generate part, the proportion is 75 %; randomly generated part, the proportion is 25 %. The maximum depth of brackets in the string corresponding to the chromosome is two layers. Randomly generated initial population will increase the diversity of the population, but will also lead to lower population fitness. In the initialization of the population, this article uses a predefined individual generation method. The predefined individuals in initial population are generated by changing each character in benchmark production, including four operations: insert a character; delete a character; replace a character; maintain the character unchanged. Using the above two methods to generate the two sub-populations, merge the two sub-populations as the initial population of the genetic algorithm.

### ***17.3.3 Genetic Operators***

The selection operator uses the method of random selection without replacement and elitist strategy. This method is to select a certain number of individuals from the population, in which choose the best individual into offspring populations. Repeat the operations until the new population size reaches the same size as the original population. This method can guarantee the best fitness individual and some individuals which fitness are higher than the average fitness is saved in the population of the middle of the process.

The cross strategy we used in this paper is one-point crossover. The crossover operator will lead to the traditional chromosome brackets illegal match. In this section, for the character sub-chromosome in a randomized crossover point for switching operations and for the bracket sub-chromosome exchange the bracket module. In this step, although it will not cause the brackets do not match, will still generate meaningless brackets, for example “[−]”.

In this paper, the mutation operator changes the number of iteration and chromosome. In the string chromosome, mutation operator only changes the single character. In the bracket chromosome will change the two brackets. As the crossover operator, mutation operator will not generate the brackets do not match, but will generate meaningless brackets.

### 17.3.4 Repair Mechanism

In this paper, we use crossover and mutation operators transform the string sub-chromosome and the brackets sub-chromosome. Although it can avoid brackets do not match, but will still generate meaningless brackets. Therefore, we use the repair mechanism to delete meaningless brackets and invalid characters. The repair mechanism to correct the meaningless brackets and invalid characters is given shown in Table 17.1.

### 17.3.5 Fitness Function Design

In genetic algorithm, fitness is the only standard to evaluate individual to guide the direction of genetic evolution. The essence of fitness function is to calculate the similarity between two tree structures. Considering the characteristics of the three-dimensional tree structure, this paper presents a similarity calculation method based on the particle size of three-dimensional tree structure. The tree structure is firstly divided into thick, medium and thin three particle size by the algorithm: similar in outline, topology and internal node. Surrounding the tree structure with the parabolic and it will reflect the different of the tree's outlines. Then get the similarity of the tree's topology can be calculated by comparing the relationship of the nodes in the tree. After that, it will gain the similarity of the internal space can be calculated by calculating the minimum space distance based on the topology relation. Finally, the similarity between the two tree structures can be calculated by combining the outline similarity, the topological similarity and the internal space similarity, as shown in Formula 17.1.

**Table 17.1** Illegal style and relative repair mechanism

Illegal style	Examples	Repair mechanism
Empty brackets or only direction control symbols	[], [-/∧]	Delete the substring with the bracket border.
Two adjacent control characters, the control characters in opposite directions offset each other	- + F	Remove the offset control characters
A plurality of consecutive identical control characters	&&&& F	Same control characters can only have four successively adjacent
The end of the string is a control character	Of F&	Remove the control characters at the end of the string
The end of the string in the brackets is a control character	[F&]	Remove the control characters at the end of the string

$$\text{similarity}(T_1, T_2) = \alpha * \text{similar\_global}(T_1, T_2) + \beta * \text{similar\_inner}(T_1, T_2) + \gamma * \text{similar\_topology}(T_1, T_2) \quad (17.1)$$

$$\text{similar\_global}(T_1, T_2) = \frac{1 - |d1 - d2|}{d1 + d2} \quad (17.2)$$

$$\begin{aligned} \text{similar\_inner}(T_1, T_2) = & \frac{1}{3} * \left( 1 - \frac{\text{total\_dist\_x}}{\text{total\_x}} \right) + \frac{1}{3} * \left( 1 - \frac{\text{total\_dist\_y}}{\text{total\_y}} \right) \\ & + \frac{1}{3} * \left( 1 - \frac{\text{total\_dist\_z}}{\text{total\_z}} \right) \end{aligned} \quad (17.3)$$

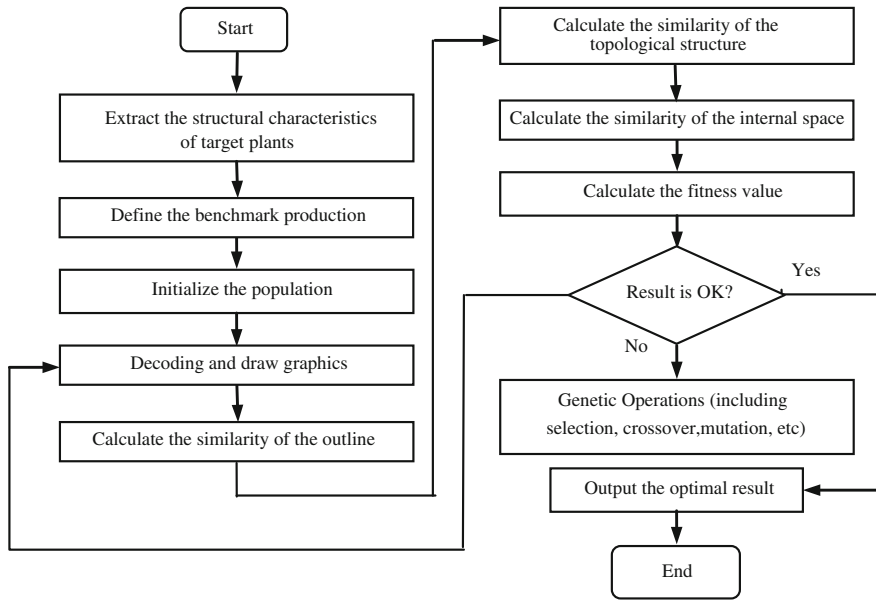
$$\text{similar\_topology}(T_1, T_2) = \frac{\text{dist\_child}}{\text{numChild\_T1} + \text{num\_Child\_T2}} \quad (17.4)$$

Here,  $\alpha$ ,  $\beta$  and  $\gamma$  are correction coefficient, and satisfy  $\alpha + \beta + \gamma = 1$ . In the function  $\text{Similar\_global}(T_1, T_2)$ ,  $d1$  and  $d2$  respectively represent the parabola focal length corresponding to the two graphical tree. In function  $\text{Similar\_inner}(T_1, T_2)$   $\text{total\_dist\_x}$ ,  $\text{total\_dist\_y}$ ,  $\text{total\_dist\_z}$ ,  $\text{total\_x}$ ,  $\text{total\_y}$ ,  $\text{total\_z}$  respectively represent the cumulative distance and cumulative value of the  $T_1$  and  $T_2$  between the nodes in the X, Y, Z direction.  $\text{dist\_child}$ ,  $\text{numChild\_T1}$ ,  $\text{numChild\_T2}$  in function  $\text{Similar\_topology}(T_1, T_2)$  respectively represent the difference between the number of a node in the same topological location of  $T_1$  and  $T_2$  and the accumulated valued of the number of node children.

## 17.4 Experiments and Results

The algorithms are written in c++ language with the OpenGL library, and it is implemented on a PC with Pentium IV 1.83 GHz CPU, 1.5 GB RAM. The detailed process of our algorithm is shown in Fig. 17.1.

Here we simulate the topological structure of a tree described in Fig. 17.2 in order to test and verify the proposed algorithm. The parameters used in the algorithm are shown in Table 17.2. In the experiment, we can measure the morphological parameters of the target tree. And then, our algorithm is used to evolve the initial population. After one hundred times iteration, we can get an excellent individual, as shown in Fig. 17.3. Ordinarily, optimal individuals can be obtained after one hundred times' iteration. The experiment results show that our method can not only simulate the special plant shape, but also simulate various plant shapes.



**Fig. 17.1** The detailed process to automatically obtain L-system rules

**Fig. 17.2** Target tree to be simulated





**Table 17.2** Parameters used in the algorithm

Items	Values
Population size	30
Termination conditions	Fitness value $\geq 0.85$
Maximum evolutionary times	100
Maximum length of chromosome	Variable
The maximum number of character “F” in a chromosome	20
The number of iterations	Variable
Mutation rate for the number of iterations	0.01
1-point crossover rate	0.8
Mutation rate for string sub-chromosome	0.01
Mutation rate for bracket sub-chromosome	0.005
Character selected rate	0.4
Rate of removing the brackets	0.3
Rate of adding brackets brackets	0.4
Rate of moving brackets	0.3

**Fig. 17.3** Simulated result of a 3D tree

## 17.5 Conclusions

To simplify the modeling processes and improve the simulating speed when we employ L-systems to describe the topological structure, we present here an intelligent algorithm to simulate the images of virtual plant based on Genetic Algorithm and L-systems. It has been proved in practice that our method is feasible and valid. In the future studies, we will discuss the following topics: (1) how to increase the amount of individuals which are accord with L-system; (2) how to

increase the variety of individuals under the condition that the individuals suit to the rule of L-system; (3) How to compare the similarity of trees is the premise of designing individual fitness function.

**Acknowledgments** This research program has been supported by National Natural Science Foundation of China under Grant No. 60901081, and Public Welfare Project of Science and Technology Department of Zhejiang Province under Grant No. 2012C32003.

## References

1. Guo Y, Fourcaud T, Jaeger M et al (2011) Plant growth and architectural modelling and its applications. *Ann Bot* 107(5):723–727 (in chinese)
2. Fourcaud T, Zhang X, Stokes A et al (2008) Plant growth modelling and applications: the increasing importance of plant architecture in growth models. *Ann Bot* 101(8):1053–1063
3. Livny Y, Pirk S, Cheng Z et al (2011) Texture-lobes for tree modeling. *Trans Graph* 30(4): Article 53
4. Vos J, Evers JB, Buck-Sorlin GH et al (2010) Functional-structural plant modelling: a new versatile tool in crop science. *J Exp Bot* 61(8):2101–2115
5. Wang F, Kang M, Lu Q et al (2011) A stochastic model of tree architecture and biomass partitioning: application to Mongolian Scots pines. *Ann Bot* 107(5):781–792
6. Sarlikioti V, de Visser PHB, Marcelis LFM (2011) Exploring the spatial distribution of light interception and photosynthesis of canopies by means of a functional-structural plant model. *Ann Bot* 107(5):875–883
7. Buck-Sorlin G, PHB de Visser, Henke M et al (2011) Towards a functional-structural plant model of cut-rose: simulation of light environment, light absorption, photosynthesis and interference with the plant structure. *Ann Bot* 108(6):1121–1134
8. Oliapuram NJ, Kumar S (2010) Real time forest animation in wind. In: Proceedings of the seventh Indian conference on computer vision, graphics and image processing, p1.22:197–204
9. Prusinkiewicz P, Lindenmayer A (1990) The algorithmic beauty of plants. Springer, Berlin
10. Jacob C (1994) Genetic L-system programming. PPSN III—parallel problem solving from nature, international conference on evolutionary computation. Lecture notes in computer science, vol 866. Springer, Berlin, pp 334–343
11. Ochoa G (1998) On genetic algorithms and lindenmayer systems. parallel problem solving from nature (PPSN V). Lecture notes in computer science, vol 1498. Springer, Berlin, pp 335–344
12. Mock KJ (1998) Wildwood: the evolution of L-systems plants for virtual environments. In: International conference on evolutionary computing '98, Anchorage, Alaska, IEEE-Press
13. Hornby GS, Pollack JB (2001) Evolving L-systems to generate virtual creatures. *Comput Graph* 26(6):1041–1048
14. McCormack J (2004) Aesthetic evolution of L-systems revisited. In: Proceedings of Evo workshops 2004, LNCS, vol 3005, pp 477–488
15. Kókai G, Tóth Z, Ványi R (1999) Evolving artificial trees described by parametric L-systems. In: Proceedings of the first Canadian workshop on soft computing, Edmonton, Alberta, Canada, pp 1722–1728
16. Traxler C, Gervautz M (1996) Using genetic algorithms to improve the visual quality of fractal plants generated with CSG-PL-systems. Research report, TR-186-2-96-04, Institute of Computer Graphics and Algorithms, Vienna University of Technology
17. Bian R, Chen P, Burrage K et al (2002) Derivation of L-system models from measurements of biological branching structures using genetic algorithms. *Lect Notes Comput Sci* 2358:514–524
18. Venter J, Hardy A (2007) Generating plants with gene expression programming. In: Proceedings of afrigraph 2007, pp 159–167

# Chapter 18

## Shrinkage Common Spatial Pattern for Feature Extraction in Brain-Computer Interface

Yu Zhang, Jing Jin, Bei Wang and Xingyu Wang

**Abstract** Common spatial pattern (CSP) has been one of the most popular methods for EEG feature extraction in brain-computer interface (BCI) application. Although the CSP usually provides good discriminant features for classification, it is also known to be sensitive to overfitting and noise. This study introduces a shrinkage technique to regularize estimation of the covariance matrices in the CSP and hence a novel shrinkage CSP (SCSP) method, which could effectively alleviate the effects of small training sample size and unbalanced data on classification. The proposed SCSP is validated on feature extraction of P300 that has been widely adopted for BCI development. Classification accuracies are evaluated by using linear discriminant analysis (LDA) with experimental EEG data from seven subjects. The results indicate that the proposed SCSP extracts more effective features that yield higher classification accuracy than that by the traditional CSP.

**Keywords** Brain-computer interface (BCI) · Common spatial pattern (CSP) · Electroencephalogram (EEG) · P300 · Shrinkage

### 18.1 Introduction

A brain-computer interface (BCI) is communication technique establishing a direct connection between human brain and computer or external device without nerve and muscle function [1]. According to different types of task paradigms, the BCI system can be developed based on various specific electroencephalogram (EEG) patterns typically including P300, steady-state visual evoked potential (SSVEP)

---

Y. Zhang · J. Jin · B. Wang · X. Wang (✉)

Key Laboratory for Advanced Control and Optimization for Chemical Processes,  
Ministry of Education, East China University of Science and Technology,  
Shanghai 200237, China  
e-mail: xywang@ecust.edu.cn

and event-related (de)synchronization (ERD/ERS). One of the key problems for an improved BCI is how to extract effective features that reflect the EEG pattern actually for the subsequent classification.

Due to the volume conduction, EEG pattern recorded from the scalp usually is of a poor spatial resolution. As an effective method for feature extraction, common spatial pattern (CSP) has been widely applied to spatial filtering of various EEG patterns for BCI applications [2–5]. The CSP usually give highly discriminative features resulting in good classification performance, whereas it is also known to be overfitting when insufficient sample set or unbalanced sample set are available, which specially happens for P300 classification [5]. To overcome the drawback, this study introduces a shrinkage CSP (SCSP) method, in which the shrinkage technique [6] is adopted to regularize the covariance matrices in the CSP towards the true ones. The proposed SCSP method is validated on the P300 data recorded from seven subjects. An accuracy comparison is implemented among the linear discriminant analysis (LDA) classification without spatial filtering, with the CSP and with the proposed SCSP. The superior classification accuracy implies that the proposed SCSP is more effective to extract EEG features in the case of small sample size and unbalanced set.

## 18.2 SCSP for Feature Extraction

### 18.2.1 Common Spatial Pattern (CSP)

CSP is one of the most popular method for EEG feature extraction, which aims at learning spatial filters  $\mathbf{w}$  to maximize the variance of one class while minimize the variance of another class as the following Rayleigh quotient:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \boldsymbol{\Sigma}_1 \mathbf{w}}{\mathbf{w}^T \boldsymbol{\Sigma}_2 \mathbf{w}}, \quad (18.1)$$

where  $\boldsymbol{\Sigma}_1$  is the spatial covariance matrix for class 1 and  $\boldsymbol{\Sigma}_2$  for class 2. Assume we are given two sets of EEG data  $\mathbf{X}_{1,i} \in \mathbb{R}^{M \times K}$  and  $\mathbf{X}_{2,i} \in \mathbb{R}^{M \times K}$  ( $M$  channels  $\times$   $K$  points) from  $i$ -th trial ( $i = 1, 2, \dots, N$ ) corresponding to the two classes, and both  $\mathbf{X}_{1,i}$  and  $\mathbf{X}_{2,i}$  have been centered. The  $i$ -th covariance matrices of the class 1 and class 2 are calculated from, respectively:

$$\boldsymbol{\Sigma}_{1,i} = \mathbf{X}_{1,i} \mathbf{X}_{1,i}^T, \quad \boldsymbol{\Sigma}_{2,i} = \mathbf{X}_{2,i} \mathbf{X}_{2,i}^T. \quad (18.2)$$

The  $\boldsymbol{\Sigma}_1$  and  $\boldsymbol{\Sigma}_2$  can then be obtained by:

$$\boldsymbol{\Sigma}_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} \boldsymbol{\Sigma}_{1,i}, \quad \boldsymbol{\Sigma}_2 = \frac{1}{N_2} \sum_{i=1}^{N_2} \boldsymbol{\Sigma}_{2,i}. \quad (18.3)$$

where  $N_1$  and  $N_2$  denote the numbers of trials in classes 1 and 2, respectively. Since the  $J(\mathbf{w})$  will not change by rescaling of  $\mathbf{w}$ , the CSP optimization problem described in Eq. (18.1) can be rewritten as:

$$\max_{\mathbf{w}} \mathbf{w}^T \boldsymbol{\Sigma}_1 \mathbf{w} \quad \text{s.t. } \mathbf{w}^T \boldsymbol{\Sigma}_2 \mathbf{w} = 1. \quad (18.4)$$

The corresponding Lagrangian is:

$$L(\lambda, \mathbf{w}) = \mathbf{w}^T \boldsymbol{\Sigma}_1 \mathbf{w} - \lambda(\mathbf{w}^T \boldsymbol{\Sigma}_2 \mathbf{w} - 1). \quad (18.5)$$

By setting the derivative of  $L$  with respect to  $\mathbf{w}$  to zero, we obtain:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}} &= 2\mathbf{w}^T \boldsymbol{\Sigma}_1 - 2\lambda \mathbf{w}^T \boldsymbol{\Sigma}_2 = 0 \\ &\Rightarrow \boldsymbol{\Sigma}_1 \mathbf{w} = \lambda \boldsymbol{\Sigma}_2 \mathbf{w}. \end{aligned} \quad (18.6)$$

The eigenvectors in Eq. (18.6) can be solved from a generalized eigenvalue problem. The optimal spatial filters are then derived by retaining the eigenvectors corresponding to the largest  $D$  eigenvalues. Assume  $\tilde{\mathbf{W}} \in \mathbb{R}^{M \times D}$  is a matrix consisting of the optimal spatial filters. The features extracted by  $\tilde{\mathbf{W}}$  from EEG data  $\mathbf{X} \in \mathbb{R}^{M \times K}$  are given by:

$$\mathbf{Y} = \tilde{\mathbf{W}}^T \mathbf{X}. \quad (18.7)$$

The features  $\mathbf{Y} \in \mathbb{R}^{D \times K}$  provide the optimal discrimination between the two classes in the least-squares sense.

### 18.2.2 Shrinkage CSP (SCSP)

Although the CSP usually gives good results for EEG feature extraction, it may be overfitting to small sample size and unbalanced data. This happens particularly for P300 feature extraction. To solve such problem, this study introduces a regularized CSP based on a shrinkage technique [6] to improve the discriminability of extracted features and hence the performance of EEG classification.

Recently, regularization technique has been increasingly applied to improve covariance matrix estimation in feature extraction and classification [7, 8]. The covariance matrix corresponding to the  $i$ -th trial of class  $c$  can be typically regularized as:

$$\tilde{\boldsymbol{\Sigma}}_{c,i} = (1 - \lambda_{c,i}) \boldsymbol{\Sigma}_{c,i} + \lambda_{c,i} v_{c,i} \mathbf{I}, \quad (18.8)$$

where  $\lambda_{c,i}$  is the regularization parameter and  $v_{c,i}$  denotes the average eigenvalue  $\text{tr}(\boldsymbol{\Sigma}_{c,i})/K$  of  $\boldsymbol{\Sigma}_{c,i}$ . The regularization parameter  $\lambda_{c,i}$  is usually selected by using cross-validation. However, the cross-validation procedure is time-consuming and

may be not applicable for small sample size. Instead of cross-validation, we estimate the regularization parameter automatically and quickly according to the shrinkage technique [6] for each covariance matrix by:

$$\lambda_{c,i}^* = \frac{N_c}{(N_c - 1)^2} \frac{\sum_{m,n=1}^M \text{var}_k(z_{c,i,m,n}(k))}{\sum_{m \neq n} s_{c,i,m,n}^2 + \sum_m (s_{c,i,m,m} - v_{c,i})^2}, \quad (18.9)$$

where  $N_c$  is the number of trials in class  $c$ ,  $s_{c,i,m,n}$  denotes the element in the  $m$ -th row and  $n$ -th column of covariance matrix  $\tilde{\Sigma}_{c,i}$ , and  $z_{c,i,m,n}(k)$  is computed as:

$$z_{c,i,m,n}(k) = ((\mathbf{x}_{c,i,k})_m - (\boldsymbol{\mu}_{c,i})_m)((\mathbf{x}_{c,i,k})_n - (\boldsymbol{\mu}_{c,i})_n), \quad (18.10)$$

with  $(\mathbf{x}_{c,i,k})_m$  being the element in the  $m$ -th row and  $k$ -th column of data  $\mathbf{X}_{c,i} \in M \times K$  and  $(\boldsymbol{\mu}_{c,i})_m$  being the element in the  $m$ -th row of mean column vector computed from  $\mathbf{X}_{c,i}$ .

### 18.3 EEG Acquisition

In this study, the proposed SCSP method is validated on the classification of P300 that has been relatively more popular for BCI development.

Seven healthy subjects (S1–S7, aged from 24 to 49) with normal or corrected-to-normal vision participated in this experiment. During the experiment, the subjects were seated to a comfortable chair 60 cm from a standard 17 in. LCD monitor (1280 × 1024 screen resolution) in a shielded room.

Figure 18.1 illustrates the P300 experimental paradigm. The experimental layout contained eight arrow directions on a black screen to simulate a wheelchair

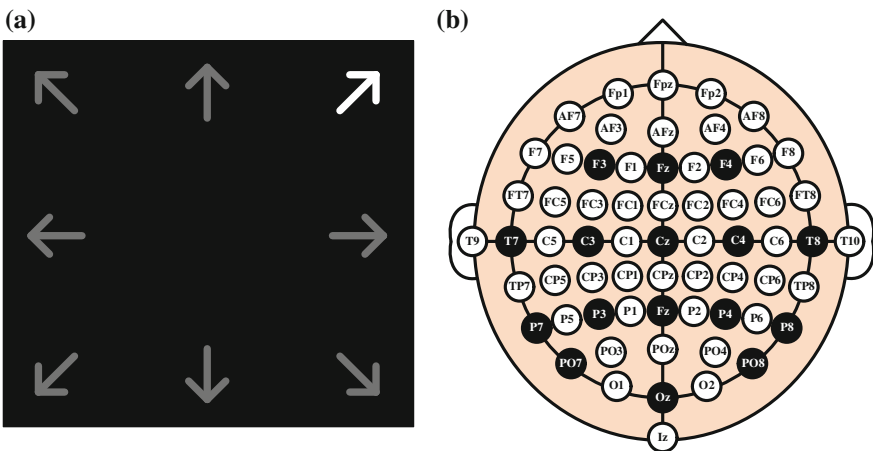
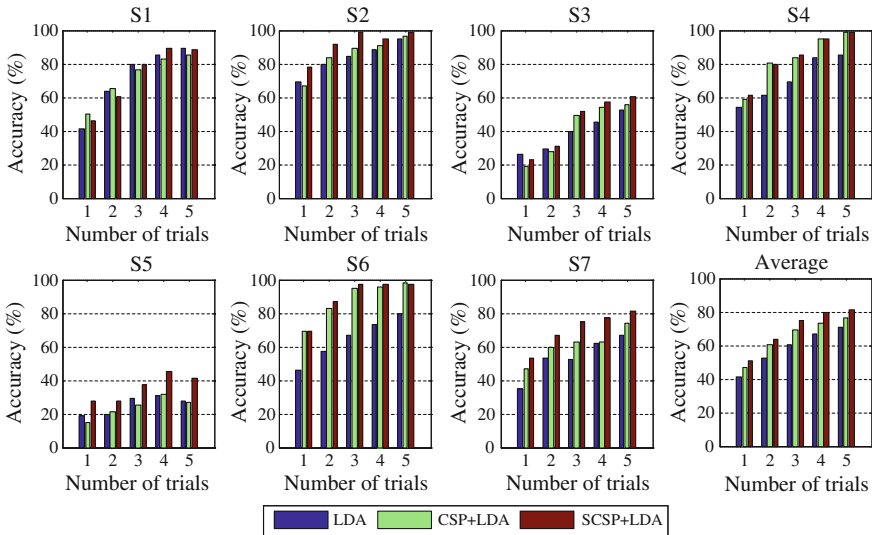


Fig. 18.1 Experimental layout (a) and channel configuration (b)

control, and intensified arrows were adopted as stimuli. Each subject completed two experimental sessions, each of which consisted of eight runs corresponding to each of the eight direction commands. In each run, a target cue was first presented for 1000 ms in the middle of the layout followed by five flash blocks. Subjects were asked to focus on the cued arrows and silently count the number of times they were intensified. In each flash block, eight arrow stimuli were intensified randomly on the corresponding eight directions for once, respectively, with 100 ms presentation duration and 80 ms inter-stimulus interval (ISI). That is, a total of 640 trials (80 targets and 560 non-targets) data were recorded from each subject.

## 18.4 Results

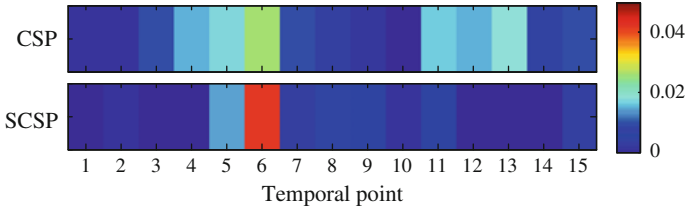
To evaluate the effectiveness of the extracted features, classification accuracy was compared among the LDA classification without spatial filtering, with the ordinary CSP and with the proposed SCSP. For each subject, half data were randomly selected from the 640 trials data for training classifier, while the remaining half for testing classifier. Figure 18.2 depicts the P300 classification accuracies obtained by the LDA method, the CSP + LDA method and the SCSP + LDA method for the seven subjects with various numbers of trials average. The results show that the SCSP + LDA method outperformed both the LDA method and CSP + LDA method for S2, S3, S5 and S7, while achieving comparable accuracies with those of the CSP + LDA method for S1, S4 and S6. Table 18.1 presents the statistical



**Fig. 18.2** Classification accuracies derived by the LDA, CSP + LDA and SCSP + LDA using various numbers of trials for the seven subjects, respectively

**Table 18.1** Statistical analysis for the classification accuracy difference among the three methods

Method comparison	Number of trials				
	1	2	3	4	5
CSP + LDA versus LDA	$p = 0.266$	$p = 0.083$	$p = 0.087$	$p = 0.104$	$p = 0.124$
SCSP + LDA versus LDA	$p < 0.05$	$p < 0.05$	$p < 0.01$	$p < 0.005$	$p < 0.01$
SCSP + LDA versus CSP + LDA	$p = 0.092$	$p = 0.11$	$p < 0.05$	$p < 0.05$	$p < 0.05$

**Fig. 18.3** Comparison of the most discriminative information ( $r^2$ -values) between the CSP and the SCSP. The 6th temporal point corresponds to 280 ms

analysis (paired  $t$  test) results of accuracy difference among the three methods. The SCSP + LDA method achieved significantly higher classification accuracy than the LDA method while the difference between the CSP + LDA method and LDA method is not significant. Furthermore, the SCSP + LDA method outperformed significantly the CSP + LDA method when using number of trials larger than two, which indicates that the regularized covariance matrices with shrinkage technique did improve the P300 classification accuracy.

A comparison of the most discriminative information was carried out between the features extracted by the CSP and the proposed SCSP. Figure 18.3 shows the most discriminative features obtained from S7 evaluated by the squared pointwise biserial correlation coefficients ( $r^2$ -values) [8]. The SCSP accurately extracted the more effective feature at 280 ms (associated to the P300 component) and prevent the disturbance from other non-associated components, compared with the traditional CSP.

The aforementioned results provide an evidence for the effectiveness of the proposed SCSP on feature extraction in the case of small sample size and unbalanced data.

## 18.5 Conclusions

This study proposed a shrinkage common spatial pattern (SCSP) method for EEG feature extraction in BCI application. The covariance matrices were regularized by the shrinkage whose parameters were automatically and quickly estimated, which



prevented overfitting to small sample size and unbalanced data. From experimental analysis with real EEG data of seven subjects, the proposed SCSP method extracted more effective features resulting in higher classification accuracy compared with the traditional CSP.

**Acknowledgments** This study was supported by the Nation Nature Science Foundation of China 61074113, 61203127, Shanghai Leading Academic Discipline Project B504, and Fundamental Research Funds for the Central Universities WH1114038.

## References

1. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM (2002) Brain-computer interfaces for communication and control. *Clin Neurophysiol* 113(6):767–791
2. Pfurtscheller G, Neuper C (2001) Motor imagery and direct brain-computer communication. *Proc IEEE* 89(7):1123–1134
3. Parini S, Maggi L, Turconi AC, Andreoni G (2009) A robust and self-paced BCI system based on a four class SSVEP paradigm: algorithms and protocols for a high-transfer-rate direct brain communication. *Comput Intell Neurosci*. Article ID 864564
4. Pires G, Nunes U, Castelo-Branco M (2011) Statistical spatial filtering for a P300-based BCI: tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis. *J Neurosci Methods* 195(2):270–281
5. Zhang Y, Zhou G, Zhao Q, Wang X, Cichocki A (2012) Regularized CSP with Fisher’s criterion to improve classification of single-trial ERPs for BCI. In: 9th international conference on fuzzy systems and knowledge discovery (FSKD), May, pp 891–895
6. Schäfer J, Strimmer K (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat Appl Genet Mol Biol* 4(1). Article 32
7. Lu H, Eng H-L, Guan C, Plataniotis K, Venetsanopoulos A (2010) Regularized common spatial pattern with aggregation for EEG classification in small-sample setting. *IEEE Trans Biomed Eng* 57(12):2936–2946
8. Blankertz B, Lemm S, Treder M, Haufe S, Müller K-R (2011) Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage* 56(2):814–825

# Chapter 19

## Detection of Overtaking Vehicles in a Highway

Song Pan and Huaping Liu

**Abstract** In this paper, some challenges of detection of overtaking cars in a highway are reviewed. Based on this analysis, we propose an overtaking vehicle detection method based on the computation of a symmetry function. Our approach defines a new symmetry function. Although the near distant vehicles usually do not show strong symmetry, we can also use this function to detect it locations. The proposed approach has been evaluated using some practical traffic scenes.

**Keywords** Vehicle detection · Computer vision · Symmetry

### 19.1 Introduction

Visual detection is one of the most challenging issues in computer vision. A lot of solutions and system realizations have been proposed which cover many applications, such as surveillance and security systems.

Currently, overtaking vehicle detection has become an important component of on-board driver assistance system. It can be used to alert the driver about driving conditions, possible collision with other vehicles, or trigger the automatic control of the vehicle for collision avoidance and mitigation. Overtaking vehicle detection based on active sensors such as laser, radar, and sonar has several drawbacks due to sensor interferences between different vehicles in a limited space. Passive

---

S. Pan

Department of Computer Software and Theory, School of Computer,  
Wuhan University, Wuhan, China  
e-mail: pans@ichaier.com

H. Liu (✉)

Department of Computer Science and Technology, Tsinghua University,  
FIT3-D02, Beijing, China  
e-mail: hpliu@mail.tsinghua.edu.cn

sensor-based detection approaches, such as vision-based methods, are becoming widely used due to their low cost and less interferences between vehicles [1].

In vision-based overtaking vehicle detection systems, a single camera is usually mounted on the host vehicle to capture rear-view image sequences. Various approaches have been proposed to detect moving vehicles assuming dynamic background. These methods can be classified into two main categories: appearance-based and motion-based. Recently, the relative motion information obtained via the calculation of optical flow becomes an important cue for detecting moving allows detection to be robust to camera shocks. Most of the corresponding research [1–7] belong to this class. However, the motion features are not oblivious when the vehicle is far from the home vehicle. In addition, optical flow approaches are very time-consuming and not very suitable for practical applications. To tackle these problems, we propose an overtaking vehicle detection method based on the computation of a symmetry function. Our approach defines a new symmetry function. Although the near distant vehicles usually do not show strong symmetry, we can also use this function to detect it locations.

The rest of the paper is as follows: The problem description is given in [Sect. 19.2](#). An overview of the proposed approach is provide in [Sect. 19.3](#). [Section 19.4](#) focuses on the detection algorithm. The Kalman filtering technology is introduced in [Sect. 19.5](#). [Section 19.6](#) gives some experimental results. Finally, [Sect. 19.7](#) gives some conclusions.

## 19.2 Problem Description

The considered scenario is constituted by a car with a camera mounted on the rear of the vehicle and looking outside. In order to exclude parts of the scene which are useless for the considered task, like the sky, the camera optical axis is pointed down; the angle made by the optical axis with the ground plane is set according to the height of the car and the maximum distance from the vehicle to be considered [4].

The system must be able to identify possible dangerous situations like a car approaching from the back or beginning an overtaking maneuver. The objective of this project is to use a camera on the side mirror to detect the passing or overtaking vehicles. Compared with detection the rear of vehicles, this project provide the following new challenges:

- (1) The head of vehicles is more complex than the rear.
- (2) Vehicle should be firstly detected in long distance. Even it is detected, there is very little feature information can be extracted, which makes the feature-based tracking very difficult. See [Figs. 19.1](#) and [19.2](#) for some examples.
- (3) The shape of the vehicle is irregular when it is near our vehicle. See [Fig. 19.3](#) for an example.

In the following we will first introduce the whole algorithm framework and then give an illustrious of the computation of the symmetry function.



**Fig. 19.1** *Left* Frame 1; *Right* frame 100



**Fig. 19.2** *Left* Frame 200; *Right* frame 240



**Fig. 19.3** *Left* Frame 244; *Right* frame 250

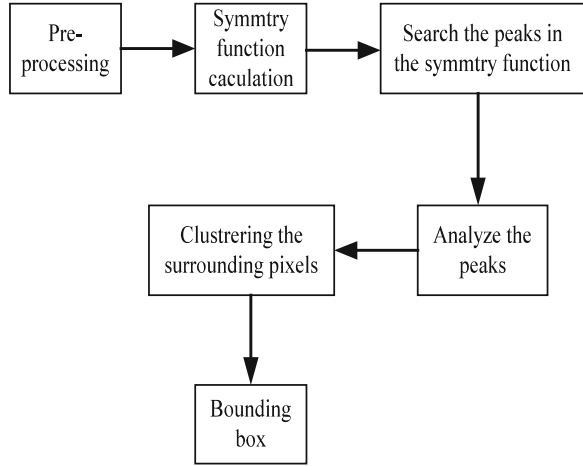
### 19.3 Detection Framework

The proposed detection framework for consists of the following four stages:

- (1) pre-processing;
- (2) symmetry function calculation;
- (3) peaks searching in the symmetry function;
- (4) pixels clustering.

The corresponding flowchart of the whole algorithm is shown in Fig. 19.4.

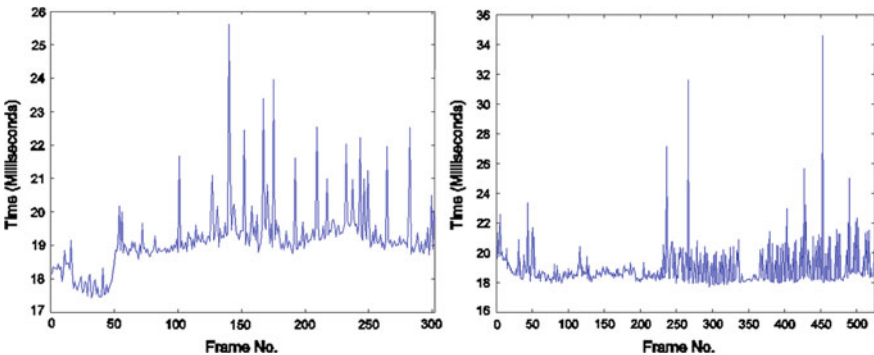
**Fig. 19.4** The detection framework



During the pre-processing stage, we binarized the image using a prescribed threshold  $th$ . This work can eliminate most of the noises (Fig. 19.5). Figures 19.6, 19.7, 19.8, 19.9 shows some of the examples. Figure 19.6 corresponds a far distance case, the right of Fig. 19.6 is the binarized image. As a comparison, we show the corresponding horizontal and vertical edge images, respectively, in Fig. 19.7. It is obvious that the binarized image is more appropriate for detection than the edge images. Also, Figs. 19.8 and 19.9 corresponds to a middle-distant vehicle.

### 19.4 Detection Algorithm

In this section, we focus on the symmetry detection. Symmetry has been studied extensively in computer vision since it is a useful principle for object detection. For the rear parts of most vehicles, the symmetry property is very common. When



**Fig. 19.5** Computation time: *Left* Sequence 1; *Right* sequence 2

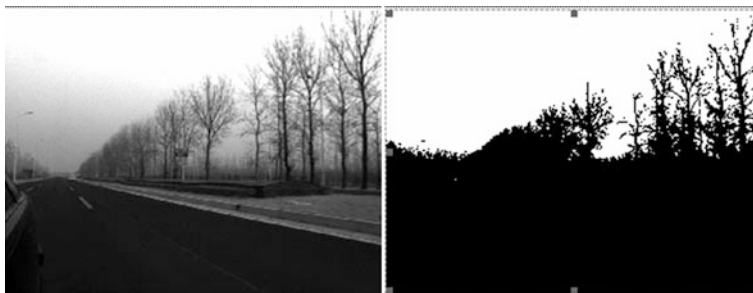


Fig. 19.6 Example of pre-processing 1

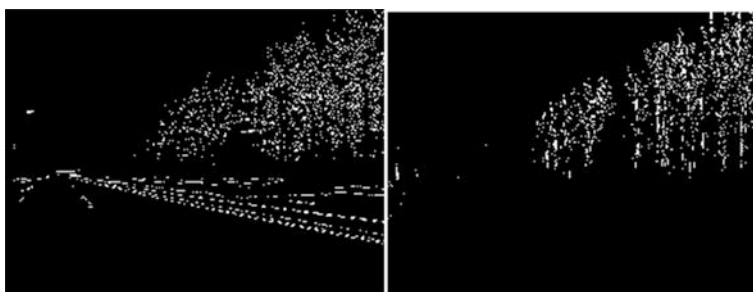


Fig. 19.7 Example of pre-processing 2

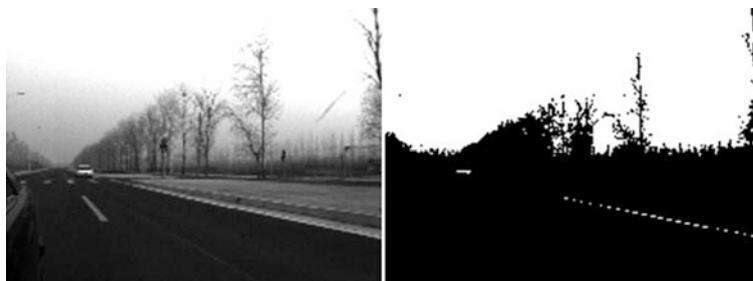


Fig. 19.8 Example of pre-processing 3

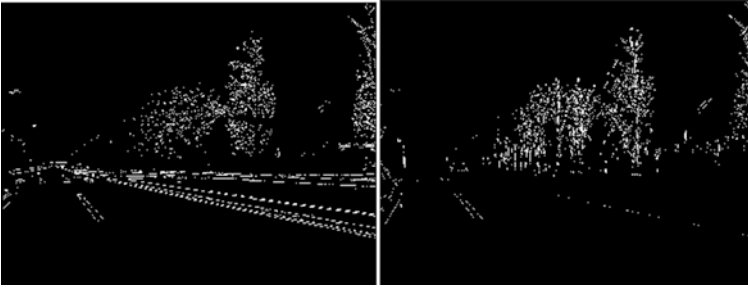
the symmetry of objects in 2-D images is considered, most researchers focus their interest and attention on the intensity symmetry and have proposed several methods regarding the detection of symmetry axes [2]. However, the symmetry will not be always obvious for the cases of overtaking vehicle detection. Therefore a carefully designed symmetry function is required. To detect the overtaking vehicle in the image, we use a window whose width is  $W$  to detect the symmetrical objects. The symmetry function  $Sym[i]$  for every column  $i$  can be obtained as follows:

**Algorithm 1:**

```

For i = W To WIDTH - W
  Sym[i] = 0;
  FOR j = 1 TO HEIGHT
    START = W;
    FOR k = START TO i
      IF ((pBinary[k,j] == 255) AND (pBinary[2*i-k,j] == 255))
        Sym[i] ++;
      END
    END
  END
END
END
END

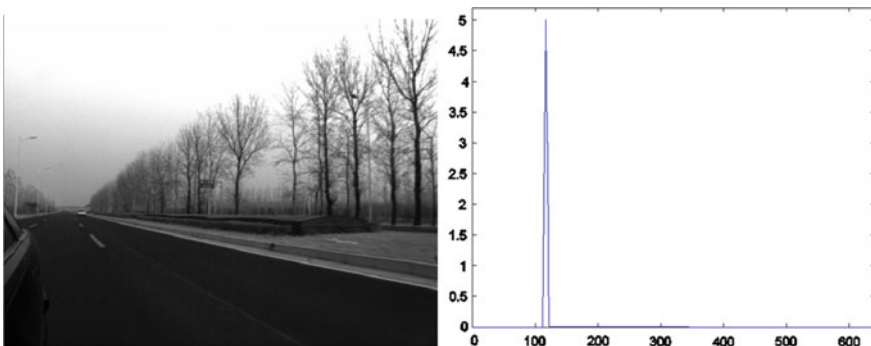
```



**Fig. 19.9** Example of pre-processing 4

where WIDTH and HEIGHT are the width and height of the considered image, respectively. pBinary represents the binarized image array.

In general, if there are vehicles in the image, there will be obvious peaks in the symmetry functions. Figures 19.10, 19.11, 19.12 and 19.13 provide some examples. From these examples, we can see that if the vehicle is far from the home



**Fig. 19.10** Frame 100

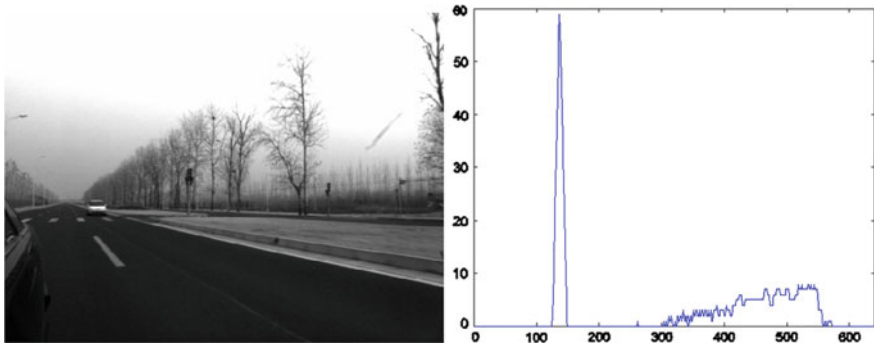


Fig. 19.11 Frame 300

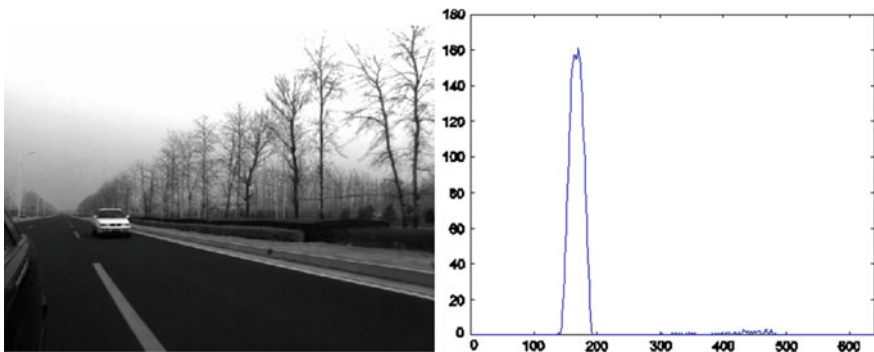


Fig. 19.12 Frame 400

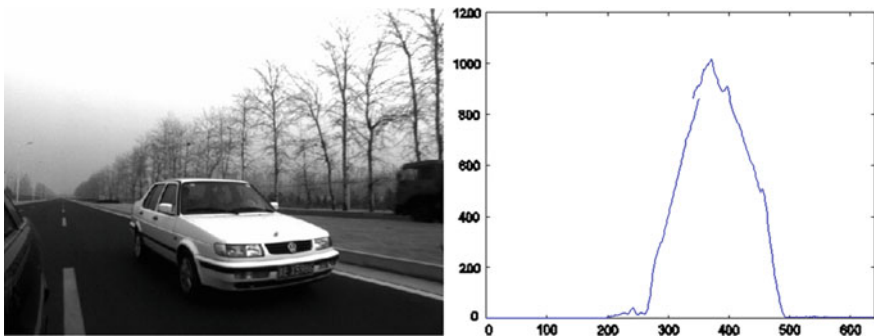


Fig. 19.13 Frame 489

vehicle, it is small and shows strong symmetry. On the other hand, when the vehicle is near the home vehicle and is overtaking, it is usually not symmetry because we can see its side. However, even in this case, the symmetry function has



an obvious peak in the neighborhood of the location of the vehicle. Therefore we can still use the symmetry function to approximately locate the vehicle.

After obtaining the symmetry function  $Sym[i]$ , we can search its peaks to locate the position of the vehicle. The following step is to obtain the exact detection bounding box. To this end, we need to cluster the surrounding pixels near the symmetrical axis to form the bounding box. In this stage, we again use the symmetry information. That is to say, only those pixels which satisfy the symmetry property are clustered.

## 19.5 Kalman Filtering

To obtain more exact results, we used the Kalman filter to refine the detection results. To this end, we construct the following state space model

$$x[k+1] = Ax[k] + Bw[k] \quad (19.1)$$

where  $x[k]$  is the state vector which contains the height, width of the detection rectangle, and  $w[k]$  is the disturbance. The measurement equation is

$$z[k] = Cx[k] + Dv[k] \quad (19.2)$$

According to the Kalman filter theory, we can get the following recursive algorithm

$$\begin{aligned} \bar{x}[k+1] &= A\hat{x}[k] \\ \bar{P}[k+1] &= AP[k]A^T + Q \\ K[k+1] &= \bar{P}[k+1]H^T(K\bar{P}[k+1]H^T + R)^{-1} \\ \hat{x}[k+1] &= \bar{x}[k+1] + K[k+1](z[k+1] - C\bar{x}[k+1]) \\ P[k+1] &= (I - K[k+1]C)\bar{P}[k] \end{aligned} \quad (19.3)$$

where  $Q$  and  $R$  are the covariance matrices of the noises  $w[k]$  and  $v[k]$ , respectively. For more details on Kalman filter, please refer to [8].

## 19.6 Experimental Results

We fixed a video camera on the left window of the host vehicle to capture the rear-view image sequences. The proposed algorithm has been implemented on a IBM PC T42 and tested on image sequence acquired from a car on different situations. The video was digitized using a sample rate of 30 frames per second. The size of each frame is 640\*480. The digitized image sequences contain various cases. Though we made a lot of experiments, due to the page limitation, we can only



**Fig. 19.14** Detection results of frame 92, 100, 200 and 280



**Fig. 19.15** Detection results of frame 300, 400, 489 and 495

include very few of them in this paper. Here, we report the detection results for two sequences. Sequence 1 contains 302 images and Sequence 2 contains 525 image. The commutation time costs are plotted in Fig. 19.5. From them we can see that the computation time is always between 17 and 35 ms and can satisfy the real time requirements.

The detection results are also rather satisfied, here we list some detection results in Figs. 19.14 and 19.15. Further, we can see the detection algorithm is rather robust to the location of the vehicle and the lane marks.

## 19.7 Conclusions

In this paper, the problem of detection of approaching car is addressed using a novel approach which is based on the computation of symmetry information. The proposed approach has several important and interesting features:

- (1) When the object is far from the home vehicle, the symmetry is very strong and therefore the detection results are rather satisfied;
- (2) When the object is approaching or overtaking, it is near the home vehicle. In this case, the symmetry is not strong. However, using our proposed symmetry function, we can also find a peak in the corresponding position. So we can locate the approximated position of the vehicle. Further, by properly clustering the surrounding pixels which satisfying the symmetry property, we can get satisfactory detection box.

In summary, the superior performance achieved in our experiments including difficult cases demonstrates the power of the new approach. These techniques also apply to effective integration of different modules of a practical vision system.

**Acknowledgments** This work is jointly supported by the National Key Project for Basic Research of China (2013CB329403), the National Natural Science Foundation of China (Grants No: 61075027, 91120011, 61210013) and the Tsinghua Selfinnovation Project (Grant No: 20111081111).

## References

1. Wang J, Bebis G, Miller R (2005) Overtaking vehicle detection using dynamic and quasi-static background modeling. In: Proceedings of IEEE computer society conference on computer vision and pattern recognition, pp 64–71
2. Du Y, Papanikolopoulos NP (1997) Real-time vehicle following through a novel symmetry-based approach. In: Proceedings of IEEE international conference on robotics and automation, pp 3160–3165
3. Zhu Y, Comaiciu D, Pellkofer M, Koehler T (2004) Passing vehicle detection from dynamic background using robust information fusion. In: Proceedings of IEEE intelligent transportation systems conference, pp 564–569
4. Tistarelli M, Guarnotta F, Rizzieri D, Tarocchi F (1994) Application of optical flow for automated overtaking control. In: Proceedings of the second IEEE workshop on applications of computer vision, pp 105–112
5. Graefe V, Jacobs U (1991) Detection of passing vehicles by a robot car driver. In: Proceedings of IEEE/RSJ international workshop on intelligent robots and systems, pp 391–396
6. Baehring D, Simon S, Niehsen W, Stiller C (2005) Detection of close cut-in and overtaking vehicles for driver assistance based on planar parallax. In: Proceedings of IEEE intelligent vehicles symposium, pp 290–295
7. Batavia PH, Pomerleau DE, Thorpe CE (1997) Overtaking vehicle detection using implicit optical flow. In: Proceedings of IEEE international conference on intelligent transportation system, pp 729–734
8. Kalman RE (1960) A new approach to linear filtering and prediction theory. *Trans ASME J Basic Eng* 82:35–46

# Chapter 20

## A Generalized Gamma Distributed CFAR Algorithm for Layover and Shadow Detection in InSAR Images

Xianxiang Qin, Huanxin Zou, Shilin Zhou and Yun Ren

**Abstract** In this paper, a novel CFAR algorithm is proposed for layover and shadow detection in Interferometric synthetic aperture radar (InSAR) images. Firstly, the generalized gamma distribution (G $\Gamma$ D) is employed for statistical modeling of the InSAR image. Moreover, a CFAR algorithm for detecting both the layover and shadow is proposed, of which the analytical expressions of two thresholds are presented. Finally, the effectiveness of the G $\Gamma$ D and the proposed CFAR algorithm are validated with a real InSAR image.

**Keywords** Interferometric synthetic aperture radar (InSAR) · Layover · Shadow · Constant false alarm rate (CFAR) detection · Generalized gamma distribution

### 20.1 Introduction

Interferometric synthetic aperture radar (InSAR) is an important technique which may be used for obtaining the high quality Digital Elevation Model (DEM) [1]. The terrain height can be accurately determined by exploiting the phase difference between the echoes of two antennas of the interferometer. However, Due to the existence of the layover and shadow phenomena in the InSAR images, this technique may become unavailable. In the past decades, many researches have been carried out for solving the problems from these phenomena [2–4], of which a special one is to locate the layover and shadow areas in the SAR images.

The constant false alarm rate (CFAR) detection is an important algorithm that has been widely used in many fields as targets detection, edge detection and so on [5], and may be employed for the layover and shadow detection. One of the most important factors of CFAR algorithm is to determine an accurate probability

---

X. Qin (✉) · H. Zou · S. Zhou · Y. Ren  
School of Electronic Science and Engineering, National University  
of Defense Technology, Changsha, China

density function (PDF) of the background. Recently, some studies have been carried out on the statistical modeling of the multilook SAR interferograms and InSAR coherence magnitude [6, 7]. Heretofore, however, as far as we know, less study has been carried out on the statistical modeling of clutter in the InSAR images.

The generalized gamma distribution (G $\Gamma$ D) is a versatile empirical distribution that has been successfully employed for statistical modeling of SAR images [8]. In this paper, we try to describe the statistics characteristics of clutter in the InSAR image with the G $\Gamma$ D. A CFAR algorithm for detecting both layover and shadow areas in InSAR images is designed. To evaluate the performance of the proposed algorithm, a real InSAR image from the TerraSAR-X system is employed.

The paper is organized as follows. The layover and shadow phenomena in the InSAR images are introduced in Sect. 20.2. Section 20.3 describes the G $\Gamma$ D and its parameter estimation as well as the proposed CFAR algorithm for detecting the layover and shadow areas. The experiments are given in Sect. 20.4. We give the conclusion in Sect. 20.5.

## 20.2 Layover and Shadow in InSAR Images

With the two acquisitions of the InSAR system, noted as  $S_1 = |S_1| \exp(j\varphi_1)$  and  $S_2 = |S_2| \exp(j\varphi_2)$  respectively, the interferogram is obtained as

$$S = S_1 \cdot S_2^* = |S_1||S_2| \exp(j(\varphi_1 - \varphi_2)) \quad (20.1)$$

where the superscript “\*” is the complex conjugate operation. Generally, according to the phase difference  $\varphi_1 - \varphi_2$  as well as the parameters of orbit and imaging sensor, the height of the target can be calculated.

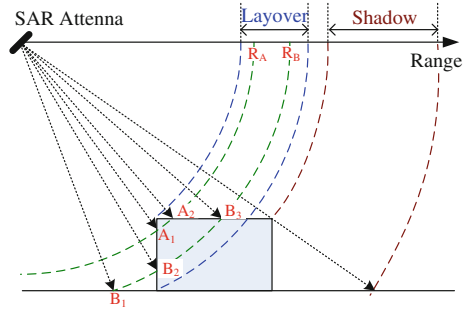
Due to the geometric relation of SAR imaging, layover and shadow are ubiquitous phenomenon in SAR images especially in areas with abrupt change like mountain and urban areas [4]. Figure 20.1 demonstrates the geometry of interferometric system in presence of layover and shadow. The layover phenomenon occurs when echoes from different patches of terrain map to the same range-azimuth resolution cell, like  $A_1$  and  $A_2$  map to  $R_A$ , or  $B_1$ ,  $B_2$  and  $B_3$  to  $R_B$  in Fig. 20.1. The shadow shows up in areas where no radar signal can reach and then only the noise can be received. Accordingly, without special procession, these phenomena would lead to unavailable information of the terrain height.

## 20.3 Generalized Gamma Distributed CFAR

### 20.3.1 G $\Gamma$ D and its Estimation

The G $\Gamma$ D was proposed by Stacy and has been widely applied in many fields [9]. A new G $\Gamma$ D which generalizes Stacy’s model is given by Li et al. [8]

**Fig. 20.1** Illustration of the geometry of interferometric system in presence of layover and shadow



$$f(x) = \frac{|v|\kappa^\kappa}{\sigma\Gamma(\kappa)} \left(\frac{x}{\sigma}\right)^{\kappa v-1} \exp\left\{-\kappa\left(\frac{x}{\sigma}\right)^v\right\}, \quad \sigma, |v|, \kappa, x > 0 \quad (20.2)$$

where  $\sigma$ ,  $v$  and  $\kappa$  are the scale, power and shape parameters respectively. This distribution is a versatile model that has many well-known distributions including the Rayleigh, Exponential, Weibull, gamma and inverse gamma distributions as special cases and the Log-normal distribution a limiting case [8].

The method of log-cumulants (MoLC) is a feasible parameter estimation for G $\Gamma$ D, which estimates parameters by solving a system of equations of log-cumulants statistics [8, 10]. The first three log-cumulants of G $\Gamma$ D can be obtained as  $\kappa_1 = \ln \sigma + (\Psi(\kappa) - \ln \kappa)/v$ ,  $\kappa_2 = \Psi(1, \kappa)/v^2$  and  $\kappa_3 = \Psi(2, \kappa)/v^3$  respectively, where  $\Psi(x) = (d/dx) \log \Gamma(x)$  and  $\Psi(n, x) = (d^n/dx^n) \Psi(x)$  refer the digamma function and the  $n$ -th order polygamma function respectively. In practice, the log-cumulants are replaced by the corresponding empirical ones [10].

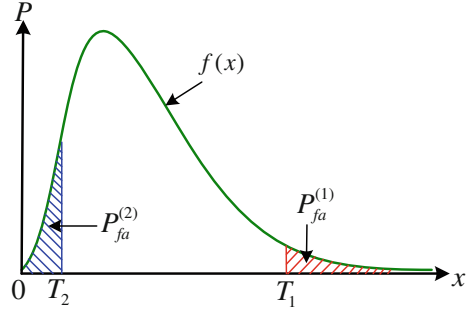
### 20.3.2 CFAR Algorithm

The CFAR algorithm performs target detection by comparing the test pixel to a threshold obtained from the statistics of the background [5, 11]. Generally, The CFAR algorithm is used for detecting “bright targets” whose amplitudes are larger than the background. For a desired false alarm rate (FAR)  $P_{fa}^{(1)}$ , the threshold  $T_1$  for detecting bright targets can be obtained from [5]

$$P_{fa}^{(1)} = \int_{T_1}^{\infty} f(x)dx = 1 - F(T_1), \quad (20.3)$$

where  $f(x)$  is the probability density function (PDF) and  $F(x)$  is the corresponding cumulative distribution function (CDF). For the case that the targets are generally darker than the clutter, with a desired FAR  $P_{fa}^{(2)}$ , the threshold  $T_2$  of CFAR algorithm can be obtained from

**Fig. 20.2** Illustration of CFAR algorithm



$$P_{fa}^{(2)} = \int_{-\infty}^{T_2} f(x)dx = F(T_2). \tag{20.4}$$

The test pixels that larger than  $T_1$  are treated as bright targets and that smaller than  $T_2$  are treated as dark targets. Figure 20.2 gives the illustration of the CFAR algorithm for detecting both bright and dark targets.

With the G $\Gamma$ D for modeling the background in the CFAR algorithm, the thresholds for detecting bright and dark targets, noted as  $T_1$  and  $T_2$ , can be derived as (20.5) [12] and (20.6), respectively.

$$T_1 = \begin{cases} \sigma \left[ (1/\kappa) \Gamma^{-1} \left( 1 - P_{fa}^{(1)}, \kappa \right) \right]^{1/v} & v > 0 \\ \sigma \left[ (1/\kappa) \Gamma^{-1} \left( P_{fa}^{(1)}, \kappa \right) \right]^{1/v} & v < 0 \end{cases}, \tag{20.5}$$

$$T_2 = \begin{cases} \sigma \left[ (1/\kappa) \Gamma^{-1} \left( P_{fa}^{(2)}, \kappa \right) \right]^{1/v} & v > 0 \\ \sigma \left[ (1/\kappa) \Gamma^{-1} \left( 1 - P_{fa}^{(2)}, \kappa \right) \right]^{1/v} & v < 0 \end{cases}, \tag{20.6}$$

where  $\Gamma^{-1}(x, a)$  is the inverse incomplete gamma function, namely the inverse function of the incomplete gamma function  $\Gamma(x, a)$  which is defined as [13]

$$\Gamma(x, a) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} \exp(-t) dt, \quad a, x > 0. \tag{20.7}$$

In the InSAR images, as layover areas correspond to signals from many different terrain patches, they are usually stronger than the background and can be seen as bright targets. However, as influenced by the speckle, some pixels of layover may be smaller than the background. Similarly, shadow areas are generally darker than the background for no signal can reach the corresponding areas. Some shadow pixels may be larger than the background due to the existence of speckle. Therefore, to reduce the influence from speckle, the morphological method [14] including close and open operations is applied. Figure 20.3 illustrates the flow chart of the proposed CFAR algorithm for the layover and shadow areas detection.

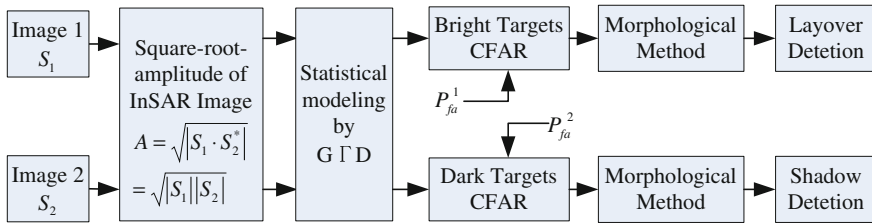


Fig. 20.3 Flow chart of the proposed CFAR algorithm

To compress the dynamic range of the interferogram amplitude, the square-root-amplitude (SRA) of InSAR image is obtained.

## 20.4 Experimental Results and Analysis

### 20.4.1 Experimental Data

In the experiments, two VV-polarized X-band amplitude TerraSAR-X images of  $1200 \times 1200$  pixels from the same mountain area of Hong Kong, noted as image 1 and image 2 and shown in Fig. 20.4a, b respectively, are employed. The resolution of the images is 1.9 m in range and 3.3 m in azimuth. The SRA of InSAR image is shown in Fig. 20.4c, which is generated according to (20.1). It can be seen from Fig. 20.4c that there are some layover and shadow areas in the mountain area.

### 20.4.2 Statistical Modeling Results for Coherent Image

The  $G\Gamma D$  is applied to model the SRA of InSAR image. For comparison, the classical Weibull [11] and  $K_A$  [8] distributions are adopted. The parameters of these distributions are estimated by the MoLC. Figure 20.5 demonstrates the

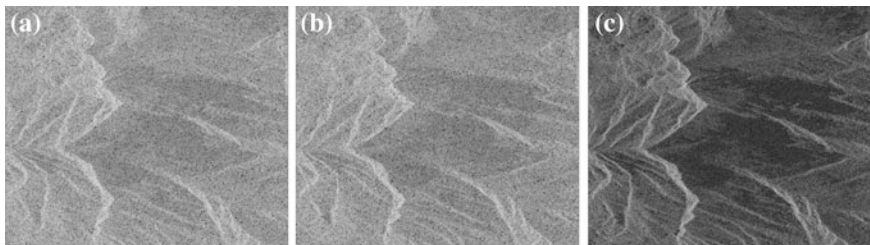
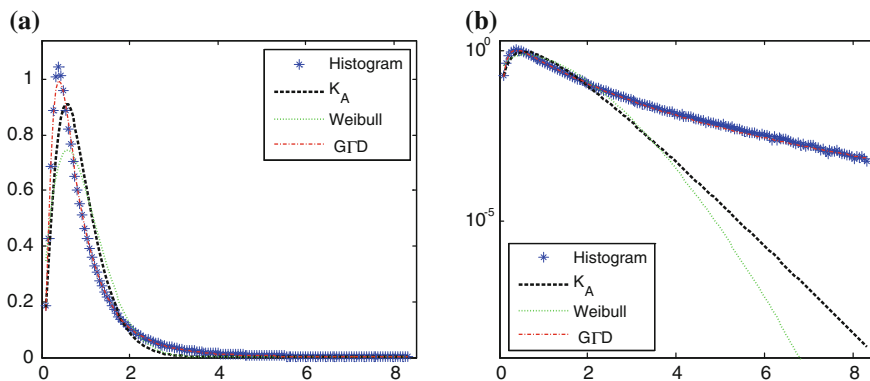


Fig. 20.4 Experimental images a image 1, b image 2, c SRA of InSAR image





**Fig. 20.5** Fitting results of estimated PDFs to histogram. **a** Fitting results in linear scale, **b** fitting results in semilogarithmic scale

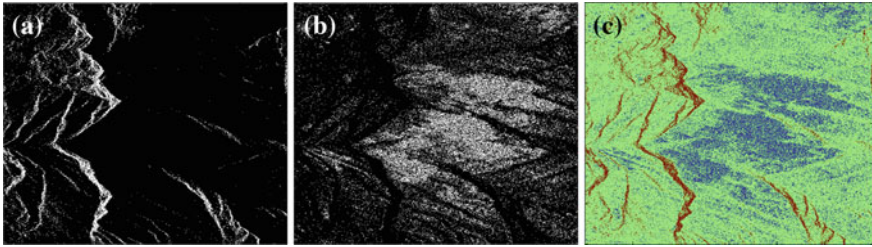
fitting results of the estimated PDFs to the histogram of the SRA of InSAR image. Figure 20.5a, b give the results in the linear and semilogarithmic scale respectively.

It can be seen that the histogram of SRA of InSAR image shows a sharp peak and a heavy tail. The  $K_A$  distribution especially the Weibull distribution diverges from the histogram obviously. In comparison, the GGD fits the histogram well. To compare these results quantitatively, the Kolmogorov-Smirnov ( $KS$ ) distance [15] is used, of which a smaller value indicates a better fitting result. The  $KS$  distances are calculated as 0.0959, 0.0745 and 0.0070 for the Weibull,  $K_A$  and GGD distribution to the histogram, respectively. This fact indicates that the GGD provides a best fitting result to the histogram compared with the two classical distributions.

### 20.4.3 Detection Results and Analysis

The desired  $P_{fa}$  is an important factor of CFAR detection algorithm, of which a larger one would lead to a higher actual detection rate while a higher actual false alarm rate simultaneously. Generally, the values of most background pixels are low, which are much closer to that of shadow pixels than to that of layover pixels. Therefore, the desired  $P_{fa}$  for detecting shadow should be larger than that for detecting layover.

In practice, in order to avoid the case that the layover or shadow areas are treated as the normal ones as possible, a large false alarm rate is usually allowed. Consequently, in the experiments, the  $P_{fa}$  for detecting the layover and shadow are given empirically as 0.05 and 0.15, respectively. The results of the layover and shadow detection according to (20.5) and (20.6) are presented in Fig. 20.6a, b respectively. The bright pixels in Fig. 20.6 indicate the positions of the layover or shadow areas. Figure 20.6c gives the detection results for both layover and shadow

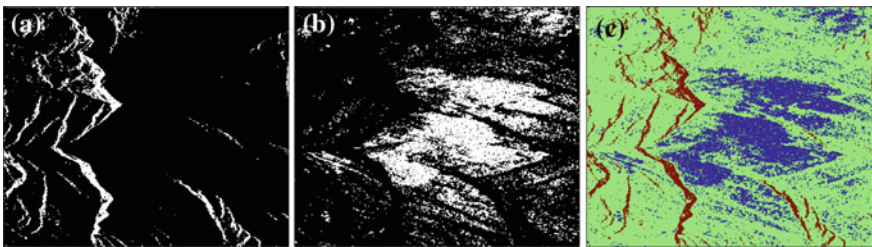


**Fig. 20.6** Layover and shadow detection by CFAR algorithm. **a** Layover detection, **b** shadow detection, **c** layover and shadow detection

areas in the same image where the brown, blue and green pixels represent the layover, shadow and the normal areas respectively.

From Figs. 20.6 and 20.4c we can see that most of the layover and shadow areas have been detected. However, there are many holes in the detected areas, so are many isolated false alarm points distributing in the images especially for the shadow detection. As analyzed before, this fact is mainly because of the influence of the speckle. To solve this problem, the morphological method of “close operation + open operation” [13] is applied on the results above. Then the obtained layover and shadow detection are shown in Fig. 20.7a, b respectively. +Figure 20.7c gives the final detection results in the same manner of Fig. 20.6c.

Comparing the results in Figs. 20.6 and 20.7, we can see that the many holes in the detected areas have been filled and numerous isolated false alarm points have been removed by the morphological method. The layover and shadow detection have been improved seriously. Generally, it is difficult to evaluate the performance of layover and shadow detection algorithm accurately due to the difficulty of obtaining precise location information of the corresponding areas. Therefore, we evaluate the proposed algorithm by comparing the original InSAR image and detection results visually. Figure 20.7 shows that the detected layover regions are mainly located at slopes faced to the SAR platform and the shadow areas locate after the mountain. These results are in accordance with the geometry of interferometric system as analyzed in Fig. 20.1. Besides, the detected results are consistent with that seen from the original image (Fig. 20.4c) too.



**Fig. 20.7** Layover and shadow detection after morphological processing. **a** Layover detection, **b** shadow detection, **c** layover and shadow detection

## 20.5 Conclusion

In this paper, we employ the  $G\Gamma D$  for the empirical statistical modeling of square-root-amplitude of InSAR image. Besides, a CFAR algorithm for layover and shadow detection with  $G\Gamma D$  for background is developed. The statistical modeling experimental result on a real InSAR image has shown the advantage of the  $G\Gamma D$  comparing with the Weibull and  $K_A$  distributions. Furthermore, the experimental results for layover and shadow detection have validated the proposed CFAR algorithm.

## References

1. Jakobsson A, Gini F, Lombardini F (2003) Layover solution in multibaseline InSAR using robust beamforming. In: Proceeding of the 3rd IEEE international symposium on signal processing and information technology, pp 328–331
2. Pairman D, McNeil S (1997) Efficient calculation in the map domain of SAR layover and shadow masks. In: 1997 IEEE international geoscience and remote sensing symposium, pp 2057–2059
3. Baselice F, Budillon A, Ferraioliand G, Pascazio V (2009) Layover solution in SAR imaging: a statistical approach. *IEEE Geosci Remote Sens Lett* 6(3):577–581
4. Cai B, Du XY, Zhen D, Liang DN (2010) Layover and shadow detection based on distributed spaceborne single-baseline InSAR. *Signal Process* 26(7):961–967 (in Chinese)
5. Gao G, Liu L, Zhao LJ, Shi GT, Kuang GY (2009) An adaptive and fast CFAR algorithm based on automatic censoring for target detection in high-resolution SAR image. *IEEE Trans Geosci Remote Sens* 47(6):1685–1697
6. Gierull CH (2004) Statistical analysis of multilook SAR interferograms for CFAR detection of ground moving targets. *IEEE Trans Geosci Remote Sens* 42(4):691–701
7. Abdelfattah R, Nicolas JM (2010) Mixture model for the segmentation of the InSAR coherence map. *Int J Appl Earth Obs Geoinf* 12S:138–144
8. Li HC, Hong W, Wu YR, Fan PZ (2011) On the empirical-statistical modeling of SAR images with generalized gamma distribution. *IEEE J Sel Topic Signal Process* 5(3):386–397
9. Stacy EW (1962) A generalization of the gamma distribution. *Ann Math Stat* 33(3):1187–1192
10. Nicolas J-M (2002) Introduction to second kind statistic: application of log-moments and log-cumulants to SAR image law analysis. *Traitement du Signal* 19(3):139–167
11. Ravid R, Levanon N (1992) Maximum-likelihood CFAR for Weibull background. *IEE Proc-F* 139(3):256–264
12. Qin XX, Zhou SL, Zou HX, Gao G (2013). A CFAR detection algorithm for generalized gamma distributed background in high-resolution SAR images. *IEEE Geosci Remote Sens Lett* 10(4):806–810
13. Abramowitz M, Stegun LA (1972) Handbook of mathematical functions. Dover, New York, p 260
14. Castleman KR (2006) Digital image processing, 2nd edn. Prentice Hall Publication, Englewood Cliffs
15. DeVore MD, O’Sullivan JA (2004) Quantitative statistical assessment of conditional models for synthetic aperture radar. *IEEE Trans Image Process* 13(2):113–125

# Chapter 21

## Error Concealment via Best Matching Block Selection Strategy

Chen Yao, Lijuan Hong and Yunfei Cheng

**Abstract** In this paper, we propose a new error concealment method based on best matching block selection strategy, which is implemented under a Bayesian probabilistic framework. Best selected image block is forwarded to a block-based bilateral filter for restoring missing pixels, while operates in block wise manner. During the Bayesian framework, a non-local spatial correlation is introduced in block selection. Meanwhile, a Markov Random Field Prior is built as an image prior model. Finally, experimental results show favorable results of proposed algorithm against traditional methods.

**Keywords** Error concealment · Block selection · Bayesian · Bilateral filter

### 21.1 Introduction

Transmission of video streams over error-prone channels is susceptible to packet loss due to congestion, fluctuation and error propagation caused by video compression. Error concealment is refers to the technique that recovers the corrupted image or video received from error-prone channel. Error concealment methods often make use of spatial and temporal correlation to alleviate the negative effect of packet loss, and can be regarded as post-processing stage of the decoder to fight against channel impairment without incurring overhead on the encoder side.

Error concealment methods are usually categorized into temporal error concealment (TEC) approaches and spatial error concealment (SEC) approaches. Basically, the TEC [1–4], methods replace corrupted block by the corresponding block in previous frame. In this case, motion vector (MV) is used for motion

---

C. Yao (✉) · L. Hong · Y. Cheng  
The Third Research Institute of Ministry of Public Security, Shanghai, China  
e-mail: yaochensing@gmail.com

compensation. Boundary matching algorithm (BMA) [5] uses the sum of absolute differences (SAD) as a criterion finding lost block MV. Except BMA algorithm, others resort to estimation (ME) for reconstructing MV [6]. The computational complexity of these TEC approaches is high because of the ME step. Therefore, some fast algorithms using fast ME or constraining search range are proposed [4], [7]. Furthermore, the MVs of neighboring blocks correctly received are interpolated instead of performing ME [8].

Spatial error concealment methods primarily utilize spatial correlation of pixels to restore missing pixels. SEC methods in the transform domain are also proposed in [9–11], an adaptive spatial interpolation is proposed in [12]. The method proposed in [13] uses spatial direction vectors extracted from edge, which is applied to recover missing pixels. The projection-onto-convex-set (POCS) [14] and fuzzy logic reasoning based methods [15] are also proposed. Furthermore, in [16], alternative projection is used in recovering the corrupted image blocks. A different SEC based on best neighborhood matching (BNM) approach is proposed in [17], which exploits block wise similarities within image to recover the missing block. The algorithm in [17] is implemented in H264 JM reference software. Unlike previous parallel methods, [18] introduced a sequential recovery mechanism together with an orientation adaptive interpolation scheme derived from pixel wise statistical model.

In this paper, inspired by forerunners in [19–22], we present a new bilateral spatial error concealment based on best block selection scheme. The image data in the corrupted region are replaced by weighted average of neighboring blocks in a local searching window. The recovery result relies on non-corrupted image data. Especially, we try to pick out a best matching block through a Bayesian rule in a local search window. Obviously, it is not a traditional weighting average scheme. We assume that isotropic region has better similarity in image. Therefore, Bayesian rule is used to select isotropic block instead of looping all the blocks in search window. After block selection, a block-based bilateral filter is implemented to recover current missing pixels.

In the rest part of this paper, the BBF and SI-BBF will be reviewed and introduced in Sect. 21.2. Abound experimental results are provided in Sect. 21.3. And finally, Sect. 21.4 concludes the paper.

## 21.2 Proposed Algorithm

In this section, the Bayesian formulation framework and block-based bilateral filter are introduced.

### 21.2.1 Bayesian Framework for Best Block Selection

Given current missing block  $P_i$  in local search window, we can get block number  $n$  through setting block size and size of search window. A non-local search window is utilized. We formulate block selection problem in a Bayesian framework as follows.

$$\begin{aligned} \arg \max_{t \in \Omega_m, k \in \{1, \dots, n-1\}, h \in \{1, \dots, m-1\}} p(P_t | P_i, P_k, P_h) &= \arg \max p(|P_i, P_k, P_h | P_t) p(P_t) / p(P_i, P_k, P_h) \\ &\propto \arg \max p(P_t | P_i) p(P_t | P_k) p(P_t | P_h) p(P_t). \end{aligned} \quad (21.1)$$

where,  $i$  is missing block index.  $t$  is matching block index.  $\Omega_m$  is matching block set.  $k$  is block index in local search window.  $h$  is block index in non-local search window.  $n$  is block number of local window.  $m$  is block number of no-local window. The Bayesian posterior probability can be factorized into block likelihoods and model prior probability.

### 21.2.2 Non-Local Likelihood Formulation

Inspired by [22], we formulate all the likelihood using non-local-mean block wise Gaussian description. All the likelihood definitions are denoted as follows,

$$\begin{aligned} p(P_t | P_i) &\propto \frac{1}{Z(i)} \exp \left\{ -\frac{\|P_i - P_t\|_{G_\sigma}^2}{h^2} \right\} \\ p(P_t | P_k) &\propto \frac{1}{Z(k)} \exp \left\{ -\frac{\|P_k - P_t\|_{G_\sigma}^2}{h^2} \right\} \\ p(P_t | P_h) &\propto \frac{1}{Z(h)} \exp \left\{ -\frac{\|P_h - P_t\|_{G_\sigma}^2}{h^2} \right\}, \end{aligned} \quad (21.2)$$

where, the parameter  $h$  controls the decay of exponential function.  $G_\sigma$  is model Gaussian variance.  $Z$  is normalizing factor. It is given by

$$\begin{aligned} Z(i) &= \sum_{t \in \Omega_m} \exp \left\{ -\frac{\|P_i - P_t\|_{G_\sigma}^2}{h^2} \right\} \\ Z(k) &= \sum_{k \in \{1, \dots, n-1\}} \exp \left\{ -\frac{\|P_k - P_t\|_{G_\sigma}^2}{h^2} \right\} \\ Z(h) &= \sum_{h \in \{1, \dots, m-1\}} \exp \left\{ -\frac{\|P_h - P_t\|_{G_\sigma}^2}{h^2} \right\}. \end{aligned} \quad (21.3)$$

We can use this non-local likelihood to measure the similarity between current block and searching block. Obviously, the searching block is more similar to current block if the computed likelihood is bigger. Here, Non-Local likelihood is used to model correlation between different blocks in the Bayesian framework. The introduction of Non-Local likelihood model is served as a key factor in the problem solving of block selection.

### 21.2.3 Prior Model

The selection of prior model is very important to the final results in a Bayesian framework. To maintain spatial and color coherence, we use a Markov Random Field (MRF) model, which asserts that the conditional probability of a pixel only depends on its neighborhood. In this paper, we use a Gaussian MRF to model image prior. This is characterized by the following local conditional probability density function:

$$p(P_t) \propto \prod_N \frac{1}{\sqrt{2\pi\sigma_c}} \exp \left\{ -\frac{(P_t - \overline{P_t(N)})^2}{\sigma_c^2} \right\}. \quad (21.4)$$

where,  $\overline{P_t(N)}$  is the mean of neighborhood of block  $P_t$ ,  $N$  denotes neighborhood block index and  $\sigma_c$  is the covariance matrix of  $P_t - \overline{P_t(N)}$ . Prior probability can be computed by the marginalization of local conditional probability.

### 21.2.4 MAP Solving

An estimate for posterior  $p(P_t|P_i, P_k, P_h)$  is found by finding the MAP estimate using the Iterated Conditional Modes (ICM) algorithm. The converged estimate represents a local optimization in the posterior formulation. Importantly, a good initialization of unknown random variables is necessary to ensure that the converged result is close to the global optimization. A multi-resolution scheme is incorporated into the algorithm. Final result is more likely to converge to the global maximum.

### 21.2.5 Block-based Bilateral Filtering

After best block selection, a block-based bilateral filter (BBF) is applied. For easy description and denotation, we use gray-scale image as an example throughout the

paper, however, error concealment algorithm of color image can be understood analogously by replacing a single image pixel with a triplets of RGB or YUV.

The classical bilateral filter is defined by

$$I'_i = \frac{1}{\sum_{j \in N_i} W_j} \sum_{j \in N_i} W_S^{(j)} \cdot W_R^{(j)} \cdot I_j, \quad (21.5)$$

Where,  $i, j$  are the pixel indexes.  $N_i$  defines a neighboring pixel index of  $i$ .  $W_j$  is normalized weight factor, which is given by

$$W_j = W_S^{(j)} \cdot W_R^{(j)}, \quad (21.6)$$

where,  $W_S = G_{\sigma_s}(\|i - j\|)$  is temporal (spatial in case of images) weights with Gaussian kernel  $G_{\sigma_s}$  and variance  $\sigma_s$ , which measures the geometric distance between  $i$  and  $j$ .  $W_R = G_{\sigma_r}(\|I_i - I_j\|)$  is radiometric weights with Gaussian kernel  $G_{\sigma_r}$  and variance  $\sigma_r$ , which measures the radiometric distance between  $I$  and  $j$ .  $I_j$  is pixel gray intensity. BBF algorithm extends  $I_j$  to gray intensity matrix of image block, which is formulated as

$$B'_k = \frac{1}{\sum_{l \in N_k} W_l} \sum_{j \in N_k} W_S^{(l)} \cdot W_R^{(l)} \cdot P_l. \quad (21.7)$$

where,  $k, l$  are block indexes.  $W_S = G_{\sigma_s}(\|k - l\|)$  and  $W_R = G_{\sigma_r}(\|B_k - P_l\|)$ . Understandably, BBF is a natural extension of bilateral filter from operating in pixel-wise manner to block-wise manner.  $B'_k$  is the final recovered block.  $P_l$  is the selected matching block.

### 21.3 Experiments

In this section, we demonstrate the error concealment results of the proposed algorithm while providing its comparisons with some other error concealment algorithm. Reconstruction results under different block loss rate are shown in Fig. 21.1. We set  $\theta$  as block loss rate with  $8 \times 8$  block size. Particularly, we consider three typical block loss rates  $\theta \approx \{25\%, 50\%, 10\%\}$ .  $\theta \approx 10\%$  represents mixed random block loss.  $\theta \approx 25\%, 50\%$  are isolated block loss. Different error patterns are shown in Fig. 21.1a, c and e respectively. Error-concealed images are shown in Fig. 21.1b, d and f respectively.

Table 21.1 illuminates the PSNR results as compared to [14, 17, 19] for Lena image under various block-loss conditions. It can be seen that we achieve highest PSNR gain over previous algorithm. Subjective results shown in Fig. 21.1 also indicate that corrupted image processed by our method has good visual quality. We attribute the improvement of PSNR to the use of best block selection scheme. The corrupted block can be well reconstructed by the most similar non-corrupted





**Fig. 21.1** Experimental results. **a** Corrupted Lena image (with 25 % block loss). **b** Recovered image of (a). **c** Corrupted Lena image (with 50 % block loss). **d** Recovered image of (c). **e** Corrupted Lena image (with mixed 10 % random block loss). **f** Recovered image of (e)

**Table 21.1** Performance comparison in PSNR

Rate of block loss	Method [17]	Method [14]	Method [19]	Ours
25 % isolated block loss	22.39	27.53	31.11	34.67
50 % isolated block loss	15.43	24.84	27.79	31.35
30 % mixed block loss	22.70	28.92	32.05	35.70

block. As compared to previous error concealment algorithm [14, 17, 19], our algorithm obtains both objective and subjective improvement. And our algorithm imposing no restrict on input image (such as block loss rate, size of lost block or no missing block on image boundary) and does not need preprocessing. And our error concealment method does not depend on parameter adjustment either.

## 21.4 Conclusion

In this paper, we present a novel error concealment algorithm built on best block selection strategy. We build a Bayesian framework through introducing non-local likelihoods and MRF prior. Finally, best selected block is forwarded into block-based bilateral filter. Experiments demonstrate the effectiveness and efficiency of block selection based error concealment algorithm.

**Acknowledgments** This work was supported by Science and Technology Innovation Foundation of Science and Technology Commission of Shanghai Municipality (12DZ0503300).

## References

1. Yao W, Stephan W, Jiangtao W (2000) Error resilient video coding techniques-real-time video communications over unreliable networks. *IEEE Signal Process Mag* 17:61–82
2. Wang Y, Zhu QF (1998) Error control and concealment for video communication: a review. *Proc IEEE* 86:974–997
3. Sun H, Zdepski JW, Kwok W, Raychaudhuri D (1997) Error concealment algorithms for robust decoding of MPEG compressed video. *Signal process Image commun* 10:249–268
4. Suh JW, Ho YS (2002) Error concealment techniques for digital TV. *IEEE Trans Broadcast* 48:299–306
5. Lam WM, Reibman AR, Liu B (1993) Recovery of lost or erroneously received motion vectors. In: 1993 IEEE international conference on acoustics, speech, and signal processing, ICASSP-93
6. Zhang J, Arnold JF, Frater MR (2000) A cell-loss concealment technique for MPEG-2 coded video. *IEEE Trans Circ Syst Video Technol* 10:659–665
7. Tsekeridou S, Pitas I (2000) MPEG-2 error concealment based on block-matching principles. *IEEE Trans Circ Syst Video Technol* 10:646–658
8. Hsu CT, Chen MJ, Liao WW, Lo SY (2005) High-performance spatial and temporal error-concealment algorithms for block-based video coding techniques. *ETRI J* 27:53–63

9. Wang, Y Zhu QF (1991) Signal loss recovery in DCT-based image and video codecs. In: Proceedings of SPIE, p 667
10. Park JW, Kim JW, Lee SU (1997) DCT coefficients recovery-based error concealment technique and its application to the MPEG-2 bit stream error. *IEEE Trans Circ Syst Video Technol* 7:845–854
11. Meisinger JA, Katrin K, Andriu K (2006) 2D frequency selective extrapolation for spatial error concealment in H. 264/AVC video coding. In: *IEEE international conference on image processing*
12. Song Li, Ma Xin (2010) Adaptive pixel interpolation for spatial error concealment. *J Signal Process Syst* 60(3):291–303
13. Kim W, Koo J, Jeong J (2006) Fine directional interpolation for spatial error concealment. *IEEE Trans Consum Electron* 52(3):1050–1056
14. Sun H, Kwok W, Center DS, Princeton NJ (1995) Concealment of damaged block transform coded images using projections onto convex sets. *IEEE Trans Image Process* 4:470–477
15. Park JW, Lee SU (1999) Recovery of corrupted image data based on the NURBS interpolation. *IEEE Trans Circ Syst Video Technol* 9:1003–1008
16. Marks JP, RJ I (2005) Recovery of image blocks using the method of alternating projections. *IEEE Trans Image Process* 14:461–474
17. Wang Z, Yu Y, Zhang D (1998) Best neighborhood matching: an information loss restoration technique for block-based image coding systems. *IEEE Trans Image Process* 7:1056–1061
18. Wang Y-K, Hannuksela M, Varsa V, Hourunranta A, Gabbouj M (2002) The error concealment feature in the H. 26L test model. In: *Proceedings of the international conference on image processing, vol, 2*, pp 729–732
19. Li X, Orchard MT (2002) Novel sequential error-concealment techniques using orientation adaptive interpolation. *IEEE Trans Circ Syst Video Technol* 12:857–864
20. Zhai G, Cai J, Lin W, Yang X, Zhang W (2008) Image error-concealment via block-based bilateral filtering. In: *2008 IEEE international conference on multimedia and expo*, pp 621–624
21. Elad M (2002) On the origin of the bilateral filter and ways to improve it. *IEEE Trans Image Process* 11:1141–1151
22. Mahmoudi M, Sapiro G (2005) Fast image and video denoising via nonlocal means of similar neighborhoods. *IEEE Signal Process Lett* 12:839–842

# Chapter 22

## Preliminary Evaluation of Classification Complexity Measures on Imbalanced Data

Yan Xing, Hao Cai, Yanguang Cai, Ole Hejlesen and Egon Toft

**Abstract** Classification complexity measures play an important role in classifier selection and are primarily designed for balanced data. Focusing on binary classification, this paper proposes a novel methodology to evaluate their validity on imbalanced data. The twelve complexity measures composed by Ho are evaluated on synthetic imbalanced data sets with various probability distributions, various boundary shapes and various data skewness. The experimental results demonstrate that most of the complexity measures are statistically changeable as data skewness varies. They need to be revised and improved for imbalanced data.

**Keywords** Data complexity · Imbalanced data · Classification

### 22.1 Introduction

Choosing a suitable and efficient classification method is critical for the success of a classification task. Since the performance of a classifier is strongly data-dependent, how to select the right classifier is a challenge [1]. Given a data set, the classical approach is the trial-and-error strategy, where diverse classifiers are tried and the one with the highest accuracy is accepted. However, the trial and error procedure is clueless, arbitrary and computationally expensive [2, 3]. An alternative solution is to estimate the complexity measures of the data and then use the measurements to guide selecting classifiers before learning [4]. Since the complexity measures describe the sources of classification difficulties, the procedure of

---

Y. Xing (✉) · Y. Cai

School of Automation, Guangdong University of Technology, Guangzhou 510006, China  
e-mail: yanxing@gdut.edu.cn

H. Cai · O. Hejlesen · E. Toft

Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

classifier selection is intellectual, goal-directed and computationally inexpensive. A variety of classification complexity measures have been proposed and successfully adopted in choosing classification algorithms for balanced data, among which those composed by Ho are the most common [2, 4].

Classification of imbalanced data is required in many domains and choosing the right classifier is more difficult [5]. For imbalanced data, the real sources of classification difficulty become ambiguous because of the skewed class distribution. To keep the validity of complexity-based classifier selection for imbalanced data, the instructive significance of the complexity measures should be statistically unchanged for various data skewness. In other words, the complexity measures should be statistically unvaried when the skewness of an imbalanced data set changes. Thus the measurements can grasp the real sources of classification difficulty and their guidance to classifier selection is significant.

Since most of the existing measures of classification complexity are primarily designed for balanced data, their effectiveness for imbalanced data should be investigated in. Focusing on binary classification, this paper proposes a methodology to evaluate the complexity measures on imbalanced data. In the following sections, Ho's complexity measures are reviewed and the limitations are indicated in Sect. 22.2. Then the evaluation methodology is proposed and several key issues are discussed in Sect. 22.3. After that, the experimental results are shown and analyzed in Sect. 22.4. Finally the work is summarized in Sect. 22.5.

## 22.2 Classification Complexity Measures

Classification complexity is defined as the degree of difficulty to learn classifiers from a data set, and is quantified with the complexity measures. The performances of classifiers are strongly data dependent, and the main sources causing classification difficulties include class ambiguity, boundary complexity, sample sparsity, feature space dimensionality, and etc. [2].

Basing on the previous work of other researchers, Ho and her colleagues composed a set of twelve complexity measures to guide classifier selection for binary classification [2, 4]. These complexity measures are listed in Table 22.1 and can be grouped into three categories [2].

- Measures of overlaps in feature values from different classes: this category consists of F1, F2, F3 and F4. The former three assess class overlap from the view of single feature dimension, while the F4 evaluates the overall overlap.
- Measures of class separability: this category includes L1, L2, N1, N2 and N3. The former two assess the linear separability of data and the rest evaluate class identifiability.
- Measures of geometry, topology, and density of manifolds: this category includes L3, N4 and T1. L3 and N4 evaluate the boundary nonlinearity, while T1 describes the class spread.

**Table 22.1** Data complexity measures for classification

Notation	Complexity measures <sup>a</sup>
F1	Maximum Fisher's discriminant ratio
F2	Volume of overlap region
F3	Maximum (individual) feature efficiency
F4	Collective feature efficiency
L1	Minimized sum of error distance by linear programming
L2	Error rate of linear classifier by LP
N1	Fraction of points on class boundary
N2	Ratio of average intra/inter class NN distance
N3	Error rate of 1NN classifier
L3	Nonlinearity of linear classifier by LP
N4	Nonlinearity of 1NN classifier
T1	Fraction of points with associated adherence subsets retained

<sup>a</sup> Formulae and algorithms used to estimate these measures can be referred in [2]

The above complexity measures have been used to guide the dynamic selection of classifiers for certain classification tasks with balanced data [6, 7]. For imbalanced data, the effectiveness of these measures is damaged and the existing work focuses on improving some of them or design new ones [8]. As far as we know, no work has been done to investigate in whether and how the complexity measures changes as the data skewness varies. To remedy the weakness, this paper proposes a novel methodology of evaluating the effectiveness of Ho's complexity measures for imbalanced data.

## 22.3 Evaluation Methodology

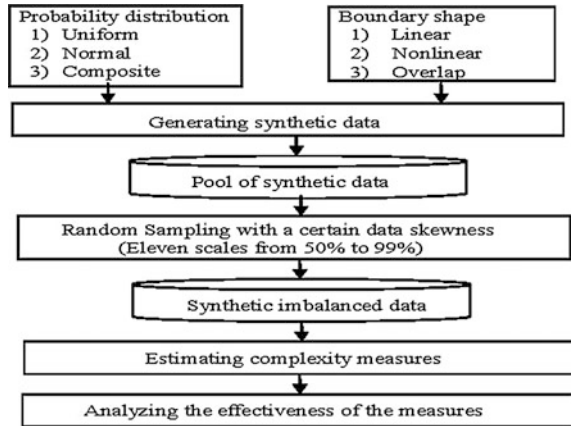
In this section, the main idea of evaluating complexity measures on imbalanced data is proposed and several key issues are discussed.

### 22.3.1 Main Idea

To obtain comprehensive and veracious results, it is required to evaluate the complexity measures on data sets with various probability distributions and boundary shapes under different data skewness. However, real data sets satisfying the above requirements are hard to be available. Therefore synthetic data are adopted. The main idea of the evaluation methodology is illustrated in Fig. 22.1.

From Fig. 22.1, it can be seen that the procedure of evaluating the complexity measures on imbalanced data consists of the following five steps.

**Fig. 22.1** Evaluation procedure



1. Configuration: set the probability distribution (uniform distribution, normal distribution or composite) and boundary shape (linear, nonlinear or overlap).
2. Generation: generate the artificial data and create a data pool.
3. Sampling: given a certain data skewness (one of the eleven scales), produce imbalanced data through random sampling from the above data pool.
4. Estimation: calculate the 12 complexity measures.
5. Analysis: analyze the effectiveness of the complexity measures through  $F$ -test.

To avoid random errors, giving a probability distribution, a boundary shape and a value of skewness, the steps of sampling and estimation are repeated twenty times and the final results are the averages.

### 22.3.2 Key Issues

In the evaluation procedure illustrated in Fig. 22.1, there are four key issues to be discussed.

1. *Probability distribution*: For the synthetic data, there are four explanatory variables ( $x_1, x_2, x_3, x_4$ ) and one binary response variable ( $y$ ). The four explanatory variables follow one of the three probability distributions listed in Table 22.2, where  $\mu$  and  $\sigma^2$  are the mean and the variance of the normal distribution respectively, while  $a$  and  $b$  are the lower and the upper bounds of the uniform distribution respectively. The response variable  $y$  is the class label and its value is decided by one of the decision functions listed in Table 22.3.
2. *Boundary shape*: Generally data can be linearly classified, nonlinearly classified or overlapped. The three kinds of boundary shapes and the involving decision functions are listed in Table 22.3.

**Table 22.2** Probability distribution of explanatory variables

Probability distribution	Formulae
Normal distribution	$X = \langle x_1, x_2, x_3, x_4 \rangle \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$
Uniform distribution	$X = \langle x_1, x_2, x_3, x_4 \rangle \stackrel{i.i.d.}{\sim} U(a, b)$
Composite distribution	$\left\{ \begin{array}{l} \langle x_1, x_3 \rangle \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2) \\ \langle x_2, x_4 \rangle \stackrel{i.i.d.}{\sim} U(a, b) \end{array} \right.$

**Table 22.3** Boundary shapes and the relating decision functions

Boundary shape	Decision function
Linear	$x_1 + 4x_2 > 0$
Nonlinear	$x_1^2 - 4x_2 > 0$
Overlap	$x_1 x_2 > 0$

3. *Data skewness*: Supposing the positive class is the rare class to be interested in, data skewness is defined as the fraction of the negative samples. The eleven scales are 50, 60, 70, 75, 80, 85, 90, 94, 96, 98 and 99 %. The sample size of the positive class decreases as the data skewness increases.
4. *Effectiveness evaluation*: To evaluate whether a complexity measure changes as the data skewness varies, the eleven values (actually each of them is the average of 20 runs) of the measure, which are relating to the eleven skewness scales, are checked to be the same or not through a single-factor *F*-test [9]. Given the significance level ( $\alpha$ ), if the values are tested to be the same, the measure is statistically unchanged as the skewness varies. Otherwise, the measure changes as the skewness varies, and then the instructive significance of the measure in classifier selection is no longer valid.

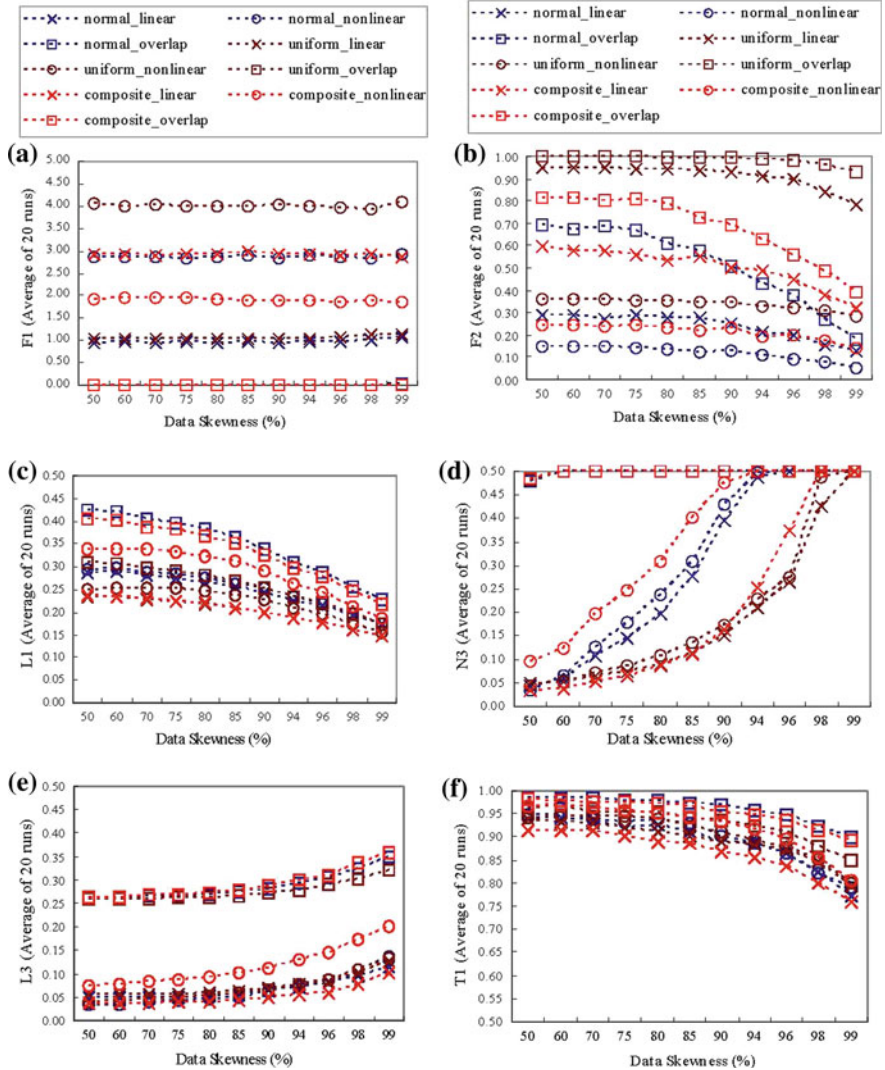
## 22.4 Evaluation Experiment

In this section, a simulation experiment is performed to evaluate the effectiveness of Ho’s complexity measures under various data skewness.

### 22.4.1 Experimental Configuration

The evaluation experiment is performed according to the procedure proposed in Sect. 22.3. In the configuration step, the normal distribution is  $N(0, 1)$  and the uniform distribution is  $U(-3, +3)$ . A data pool consists of one million samples with one of the three probability distributions and one of the three boundary





**Fig. 22.2** Classification complexity measures under various data skewness. **a** F1 under various skewness, **b** F2 under various skewness, **c** L1 under various skewness, **d** N3 under various skewness, **e** L3 under various skewness, **f** T1 under various skewness

shapes. Therefore there are nine data pools created in step 2. From one data pool,  $11 \times 20$  imbalanced data sets are produced using the eleven scales of data skewness. Thus there are 1980 unbalanced data sets produced in the sampling step, each of which consists of 10,000 samples. The algorithms of the data generation, the complexity measure estimation and the one-factor  $F$ -test ( $\alpha = 5\%$ ) are implemented with Java language, R package and the data complexity library [10].

### 22.4.2 Experimental Results

Due to the limitation of space, only some of the experimental results, such as the values of complexity measures of F1, F2, L1, N3, L3 and T1 under various data skewness, are shown in Fig. 22.2.

Looking into Fig. 22.2, it can be seen that as the data skewness increases from 50 to 99 %, the values of F2, L1 and T1 decreases, while the values of N3 and L3 increases. Only the values of F1 remain virtually unchanged.

Performing *F*-test on all the twelve complexity measures, it can be obtained that only F1 is statistically unchanged as the data skewness increases. The other eleven measures are statistically changeable as the data skewness varies. Therefore, the effectiveness of most of Ho's complexity measures becomes invalid for imbalanced data. They need to be revised and improved when being used to guide classifier selection for imbalanced data.

## 22.5 Conclusion

The classification complexity measures can grasp the sources of classification difficulties and thus play an important role in classifier selection. However, their instructive significance will be damaged by the skewed class distributions. To assess the validity of Ho's complexity measures on imbalanced data, a novel methodology is proposed to evaluate whether the complexity measures change under different data skewness. The experimental results demonstrate that most of Ho's complexity measures need to be revised and improved for imbalanced data.

The further work of this study is to evaluate the obtained conclusions on real world imbalanced data.

**Acknowledgments** This work is partially supported by the Science and Technology Project of Guangdong Province (No. 2012B050600028, No. 2012B091000171 and No. 2011B090400460) and the National Natural Science Foundation of China (No. 61074147, No. 60374062).

## References

1. Murphy KP (2012) Machine learning: a probabilistic perspective. The MIT Press, Cambridge
2. Ho TK, Basu M, Law MHC (2006) In: Basu M, Ho TK (eds) Data complexity in pattern recognition. Measures of geometrical complexity in classification problems, Springer, Berlin, pp 3–24
3. Moran S, He Y, Liu K (2009) Choosing the best bayesian classifier: an empirical study. IAENG Int J Comput Sci 36(4):9–19
4. Ho TK, Basu M (2002) Complexity measures of supervised classification problems. IEEE Trans Pattern Anal Mach Intell 24(3):298–300

5. Sun Y, Wong AC, Kamel MS (2009) Classification of Imbalanced data: a review. *Int J Pattern Recognit Artif Intell* 23(4):687–719
6. Ho TK (2008) Data complexity analysis: linkage between context and solution in classification. *Lect Notes Comput Sci* 5342:986–995
7. Ho TK (2002) A data complexity analysis of comparative advantages of decision forest constructors. *Pattern Anal Appl* 5(2):102–112
8. Weng CG, Poon J (2010) CODE: a data complexity framework for imbalanced datasets. *Lect Notes Artif Intell* 5569:16–27
9. Moore DS, McCabe GP, Craig BA (2009) *Introduction to the practice of statistics*, 6th edn. W.H. Freeman, New York
10. Orriols-Puig A, Macia N, Ho TK (2010) Documentation for the data complexity library in C++. Universitat Ramon Llull, La Salle

# Chapter 23

## A Real-Time Tracking Algorithm Based on Gray Distribution and Distance Kernel Space

Weixing Li, Yating Xiao, Feng Pan and Kai Zhou

**Abstract** The application of the traditional Camshift algorithm, which exhibits a good tracking performance in case of the obvious color characters, meanwhile, is limited in the target tracking in the color space. A fast tracking algorithm based on gray value distribution and distance kernel space is proposed. A 1.5D gray histogram method is designed to describe the model of moving object, which improves the reduction of computation for the back projection and real-time tracking performance. Moreover, a distance kernel function, describing the object weights, is constructed so as to handle the background disturbance and occlusion problem. Experiment results demonstrate the efficiency of proposed algorithm, that it can achieve a fast object tracking and resist background disturbance in some level.

**Keywords** Object · Tracking · Spatial histogram · Kernel space · Camshift

### 23.1 Introduction

Moving object tracking has been extensively used in intelligent video surveillance, intelligent transportation, human–computer interaction and virtual reality technology [1–3], and has become one of the research hotspots in pattern recognition,

---

W. Li (✉) · Y. Xiao · F. Pan · K. Zhou  
School of Automation, Beijing Institute of Technology, Beijing 100081,  
People's Republic of China  
e-mail: liweixing@bit.edu.cn

Y. Xiao  
e-mail: xiao20081654@163.com

F. Pan  
e-mail: andropanfeng@126.com

K. Zhou  
e-mail: zhokai1986@bit.edu.cn

computer technology and machine vision field. Many scholars have proposed a variety of object tracking algorithm and their related application, Funkunaga [4] proposed Meanshift algorithm, with the advantages of no parameters and fast matching, which is widely applied to the object tracking. However, this algorithm cannot automatically update the model in the process of tracking the target, which leads to tracking failure when the target size changes too much. To solve the problem, Comaniciu et al. [5] proposed Camshift algorithm, which is an application of Meanshift algorithm in continuous image sequence. The algorithm is a kind of tracking method based on HSV color space feature, firstly it finds out the position and size of the moving target in video, secondly using the position and size initializes the search window to find out the position and size of the moving target in the next frame, then repeat this process and realize the goal of continuous object tracking.

Since Camshift algorithm is based on HSV color space feature, and color feature has a strong clustering, this algorithm performs well when the target has the obvious color feature. However, the data of color feature is often in large quantities which would influence the real-time performance. At the same time, when target and background has the similar color or target occlusion, the algorithm performs not so well.

Based on Camshift algorithm, this paper proposed a fast tracking algorithm based on gray value distribution and distance kernel space. In gray space, the calculated quantities of object tracking can be greatly reduced and the real-time ability improved. The second part reviewed the process of Camshift algorithm, while the third part proposed a 1.5D histogram and the distance kernel function to solve background disturbance and occlusion problems. Lastly, by comparing to Meanshift algorithm and partial filter algorithm, we verified the feasibility of the algorithm proposed in this paper.

## 23.2 Process of Camshift Algorithm

Both Meanshift algorithm and Camshift algorithm are based on maximum gradient descent target optimization method. Camshift algorithm used the gradient of a density function, which is in adynamic variation distribution, to estimate the parameters. It always uses the HSV color space, as the color feature having a strong clustering ability, Camshift algorithm perform well in object tracking when the target owns obvious color feature. Camshift algorithm realizes through query mode, hence it has high real-time performance, small memory occupancy and suit to the embedded systems.

Camshift algorithm mainly includes the following parts. Firstly, establish the probability distribution of the target template, and calculate the back-projection view. Secondly, iteration tracking calculate by using Meanshift algorithm. Thirdly, self-adaptively adjust the tracking window. Follow is the main process of the algorithm.

1. Select the target search window in the color probability distribution drawing.
2. Calculate the zero-order moment and the centroid of the search window.
3. Move the center of search area to the centroid with the size not change, then judge whether the centroid is convergence or not, in another words, whether the change of the centroid is less than a given threshold or not.
4. Repeat step two and three until the centroid is convergence.
5. Return the spindle direction and length of the search result and output the search result.

### 23.3 Fast Tracking Algorithm Based on Gray Value Distribution and Distance Kernel Space

#### 23.3.1 Establish the 1.5D Histogram

Histogram is an important parameter in express target characteristics, which has the strong capacity of resisting disturbance of target rotation and occlusion. Meanwhile it loses the spatial information of the target. Birchfield [6] integrated histogram with image spatial information of the gray distribution and put forward second-order histogram. Due to the change of the gray distribution of the moving object region is mainly in vertical, and the distribution is uniform in horizontal, this paper purposed a 1.5D histogram, which defined as below.

$$\begin{aligned} h_u^{(1.5)}(n_u, \mu_u, \sigma_u^2) &= f(n_u, \mu_u, \sigma_u^2) \\ u &= 0, \dots, L-1 \end{aligned} \quad (23.1)$$

where,  $n_u$  is the pixel number in  $b$ th bins,  $\mu_u$  and  $\sigma_u^2$  are the mean and variance of pixel  $(x, y)$  at  $y$ -axis in  $b$ th bins,  $L$  is the number of histogram intervals. Similarity measurement at  $y$ -axis of each bin defined as below.

$$k(y, \mu_u, \sigma_u^2) = \exp \left[ -\frac{1}{2} \left( \frac{y - \mu_u}{\sigma_u} \right)^2 \right] \quad (23.2)$$

#### 23.3.2 Design the Distance Kernel Space

When target has been disturbed by similar background or occlusion, the pixels beside the target candidate region are the most vulnerable. To reduce the weight of those pixels, define the distance kernel function  $K(x, y, T)$  in line with detective target position  $T(x_c, y_c, w, h)$  as below.

$$K(x, y, T) = \exp \left\{ -\frac{1}{2} \left[ \frac{(x - x_c)^2}{(h/2)^2} + \frac{(y - y_c)^2}{(w/2)^2} \right] \right\} \quad (23.3)$$

In which,  $(x, y)$  is the pixel in target candidate region. The further  $(x, y)$  is apart from  $(x_c, y_c)$ , the smaller the value of  $K(x, y, T)$ .

### 23.3.3 Establish the Back-Projection Views

To any point  $p(x, y, T)$  in target candidate region  $H(x_c, y_c, w, h)$ , define its similarity measurement with the template as follow.

$$\begin{aligned} \text{BackMap}(x, y, u) &= \hat{q}(u)k(y, \mu_u, \sigma_u^2)K(x, y, T) \\ &= \hat{q}(u) \exp \left[ -\frac{1}{2} \left( \frac{y - \mu_u}{\sigma_u} \right)^2 \right] \exp \left\{ -\frac{1}{2} \left[ \frac{(x - x_c)^2}{(h/2)^2} + \frac{(y - y_c)^2}{(w/2)^2} \right] \right\} \\ &= \hat{q}(u) \exp \left\{ -\frac{1}{2} \left[ \frac{(x - x_c)^2}{(h/2)^2} + \frac{(y - y_c)^2}{(w/2)^2} + \frac{(y - \mu_u)^2}{\sigma_u^2} \right] \right\} \end{aligned} \quad (23.4)$$

Besides query  $\hat{q}(u)$ ,  $\text{BackMap}(x, y, u)$  has combined the distribution of the target candidate region at  $y$ -axis with the distance kernel function, which lead the weight of pixels deviating the center too much reduced.  $\text{BackMap}(x, y, u)$  is the back-projection view, where the higher the pixel value, the more likely it belong to the target (Fig. 23.1).

### 23.3.4 Determine the Target Status Information

The paper determines the target position by its first-order moment, and amends the target frame size by local maximization Bhattacharyya coefficients [7, 8].

#### 23.3.4.1 Calculate the Target Center Position

With target back-projection view, we can calculate the target first-order moment and centroid  $(x_c^1, y_c^1)$  by Eq. (23.5) and obtain a new target region  $H(x_c^1, y_c^1, w, h)$ .

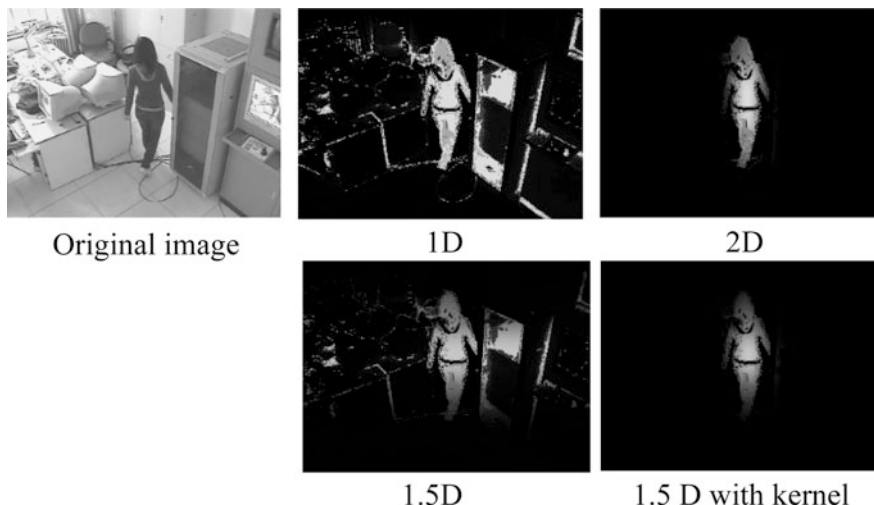


Fig. 23.1 Back-projection views of different spatial histogram

$$\begin{cases} Z_{00} = \sum_x \sum_y \text{BackMap}(x, y) \\ Z_{10} = \sum_x \sum_y x \cdot \text{BackMap}(x, y) \\ Z_{01} = \sum_x \sum_y y \cdot \text{BackMap}(x, y) \\ x_c^1 = \frac{Z_{10}}{Z_{00}}, y_c^1 = \frac{Z_{01}}{Z_{00}} \end{cases} \quad (23.5)$$

#### 23.3.4.2 Amend the Target Frame Size

Although Camshift algorithm can adaptively adjust the target frame size based on HSV space, it performs not well for gray image. To solve this problem, we amend the target frame size by local maximization Bhattacharyya coefficients.

Assume

$$H(x_c^1, y_c^1, w, h), H(x_c^1, y_c^1, w + \alpha w, h + \beta h), H(x_c^1, y_c^1, w - \alpha w, h - \beta h)$$

are the original, expanded and shrunk regions of the target. In the paper, we set  $\alpha = \beta = 0.1$ , and calculate their probability distribution  $\hat{p}^o, \hat{p}^b, \hat{p}^s$ ,  $\rho[\hat{p}^o, \hat{q}]$ ,  $\rho[\hat{p}^b, \hat{q}]$ ,  $\rho[\hat{p}^s, \hat{q}]$  as below.



$$\begin{cases} \rho[\hat{p}^o, \hat{q}] = \sum_{u=0}^{L-1} \sqrt{\hat{p}_u^o \cdot \hat{q}_u} \\ \rho[\hat{p}^b, \hat{q}] = \sum_{u=0}^{L-1} \sqrt{\hat{p}_u^b \cdot \hat{q}_u} \\ \rho[\hat{p}^s, \hat{q}] = \sum_{u=0}^{L-1} \sqrt{\hat{p}_u^s \cdot \hat{q}_u} \end{cases} \quad (23.6)$$

Thus, we obtain the update strategy of  $(w, h)$  as below.

$$(w, h) = \begin{cases} (w + \alpha w, h + \beta h) & \text{if } \rho[\hat{p}^o, \hat{q}] > (1 - \lambda) \cdot \rho[\hat{p}^o, \hat{q}] \\ (w - \alpha w, h - \beta h) & \text{if } \rho[\hat{p}^o, \hat{q}] > (1 + \lambda) \cdot \rho[\hat{p}^o, \hat{q}] \\ (w, h) & \text{else} \end{cases} \quad (23.7)$$

Besides, to prevent the divergence of the target window in actual running, we restrict the range of target window size  $(w, h)$  as below.

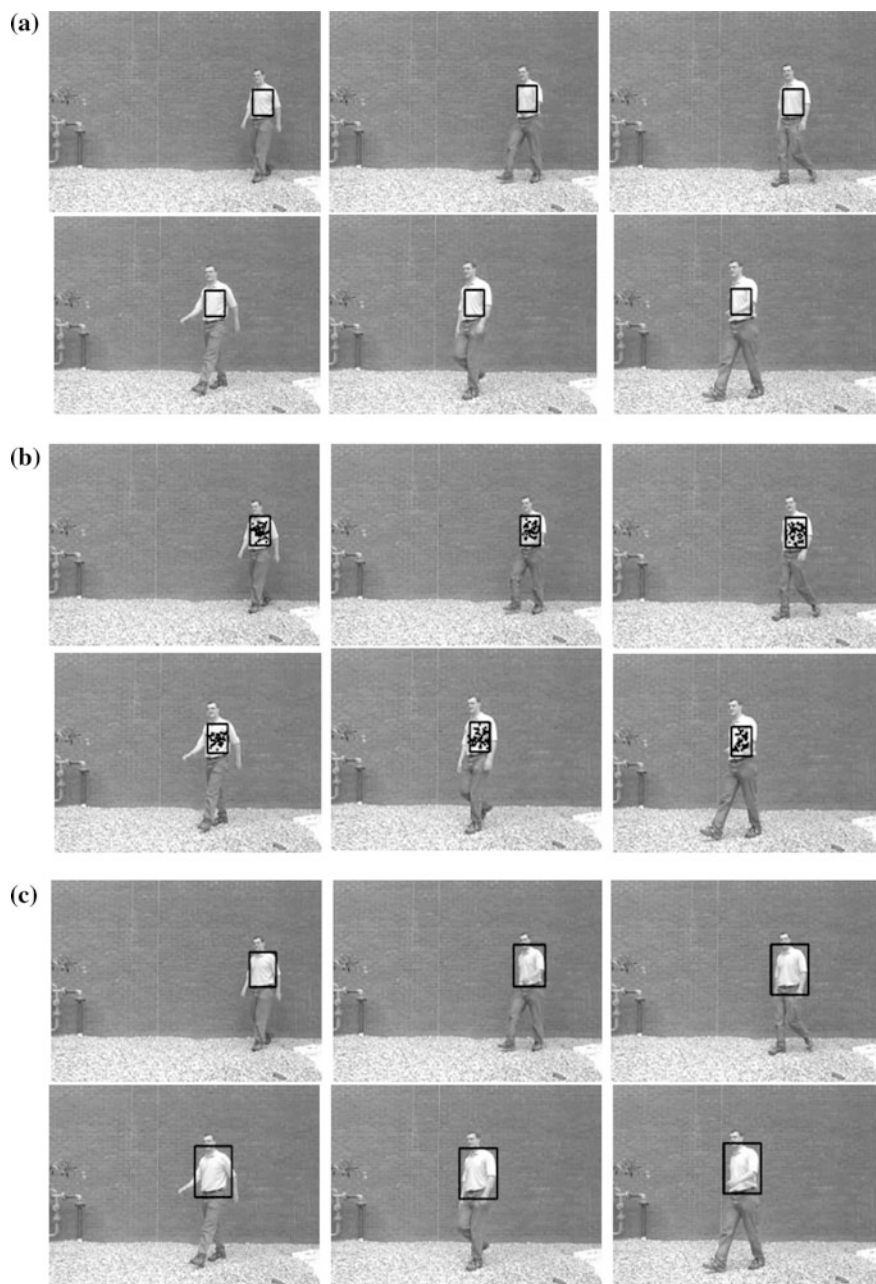
$$w = \begin{cases} w_o & \text{if } \frac{w}{w_o} > (1 + \eta) \cdot \frac{w}{w_o} < (1 - \eta) \\ w & \text{else} \end{cases} \quad (23.8)$$

where  $\lambda, \eta \in [0, 1]$ , and we set  $\lambda = 0.2, \eta = 0.2$ .

## 23.4 Experimental Results

We compared the performance of those tracking algorithms by test results on the hardware platform of Intel(R) Core(TM) 2 CPU T5500 @ 1.66 GHz processor with 1 GB memory where videos were captured at 25 frames/s and  $320 \times 240$  resolution. The test mainly tracked the upper part of the objects. Figure 23.2a is the results of Meanshift algorithm based on gray image, and the maximum iterations  $I = 10$ . Figure 23.2b is the results of particle filter based on gray image, and the particle number  $N = 10$ . Figure 23.2c is the results of the fast tracking algorithm proposed in this paper. Table 23.1 lists the runtimes of those three algorithms at the same test samples.

From the experiments results, the proposed method tracked the target through the look-up table, which not only has higher real-time ability than other two algorithms but also can adaptively adjust the target frame size based on the target movement. In conclusion, the fast tracking algorithm based on gray distribution and distance kernel space proposed in the paper can successfully track the target meanwhile consider the real-time ability of the system in the condition of good target gray statistical characteristics.



**Fig. 23.2** Tracking results of three algorithm from *left to right*, the frame numbers are 1, 11, 21, 31, 41, and 51, respectively. **a** The result of Meanshift algorithm based on *gray* image, **b** the result of particle filter based on *gray* image, **c** the result of the fast tracking algorithm in this paper proposed

**Table 23.1** Comparison of those three algorithms real-time ability

Algorithm	Average computation time for each frame
Meanshift	20.451
Particle filter	54.062
Fast tracking proposed in the paper	5.126

## 23.5 Conclusions

Based on Camshift algorithm, the paper proposed a 1.5D histogram to express target gray features, and it solve the problem of Camshift algorithm low tracking precision in gray space. To reduce the influence of background disturbance and occlusion to target, we purposed distance kernel function. The experiments results indicate that our algorithm has a higher real-time ability than some common used algorithms.

**Acknowledgments** This work is partly supported by theNatural Science Foundation (Grant No.60903005).

## References

1. Girisha R, Murali S (2010) Object segmentation from surveillance video sequences. In: First International Conference on Integrated Intelligent computing (ICIIC), Bangalore, India, pp 146–153
2. Huang L, Barth M (2010) Real-time multi-vehicle tracking based on feature detection and color probability model. 2010 IEEE intelligent vehicles symposium, San Diego, CA, USA, pp 981–986
3. Bradski GR (1988) Real time face and object tracking as a component of a perceptual user interface. In: Proceedings of the fourth IEEE workshop applications of computer vision, Berlin, Germany, pp 214–219
4. Fukunaga K, Hostetler LD (1975) The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans Inf Theory* 21(1):32–40
5. Comaniciu D, Ramesh V, Meer P (2000) Real-time tracking of non-rigid objects using mean shift. In: Proceedings of the IEEE conference on computer vision and pattern recognition, vol 2, pp 142–149
6. Birchfield ST, Sriram R (2005) Spatiograms versus histogram for region-based tracking. In: IEEE conference on computer vision and pattern recognition (CVPR), San Diego, CA, pp 1158–1163
7. Kailath T (1967) The divergence and Bhattacharyya distance measures in signal selection. *IEEE Trans Commun Technol* 15(1):523–531
8. Li L, Feng Z, Chen W et al (2008) A Coarse-to-fine searching method with kernel matching based on Bhattacharyya coefficients. *Pattern Recognit Artif Intell* 21(4), 514–519 (in Chinese)

# Chapter 24

## Laplacian Regularized D-Optimal Design for Remote Sensing Image Classification

Kang Liu and Xu Qian

**Abstract** Obtaining training sample for remote sensing image classification is time consuming and expensive especially for relatively inaccessible locations. Therefore, determining which unlabeled samples would be the most informative if they were labeled and used as training samples is the most delicate phase. Particularly, we consider the problem of active learning in remote sensing image classification. However, Classical optimal experimental design approaches are based on least square errors over the labeled samples only. They fail to take into account the unlabeled samples. In this paper, a manifold learning technique which is performed in the sample space by using graph Laplacian is applied to reflect the underlying geometry of the sample. By minimizing the least square error with respect to the optimal classifier, we can select the most representative and discriminative sample for labeling. The effectiveness of the proposed method is evaluated by comparing it with other active learning techniques existing in the literature. Experimental results on data set confirmed the effectiveness of the proposed technique.

**Keywords** Active learning · Graph laplacian · Hyperspectral imagery · Machine learning

### 24.1 Introduction

Recent advances in remote sensing technology have made remote sensing image, such as hyperspectral imagery, with hundreds of narrow contiguous bands more widely available. Thus, the image can reveal subtle differences in the spectral of

---

K. Liu (✉) · X. Qian  
School of Mechanical Electronic and Information Engineering, China University of Mining and Technology, Beijing, Room 707A, Science and Technology Building, School Road No. 11, Haidian District, Beijing 100083, China  
e-mail: liukang1112@gmail.com

land cover classes which appear similar when viewed by previous sensors [1]. Because of the very high dimensionality of the image, supervised methods need to collect a large set of labeled samples to estimate parameters for classifiers. But, it is very expensive and time consuming to obtain labeled samples. Therefore, defining an efficient training set is the most delicate phase for the success of remote sensing image classification. There is need for a procedure that builds training set as small as possible and selects samples that are important for the quality of the prediction. This kind of procedure is known in the machine learning literature as active learning. In the active learning framework, it aims at building efficient training sets by iteratively improving the performance through sampling. Active learning becomes effective when confronted to problems dealing with large amounts of sample. Nonetheless, applications in remote sensing are rare [2].

The focus of this paper is on remote sensing image classification using few labeled samples. There are two popular groups of active learning algorithms in machine learning literature. One group of algorithms select the most uncertain sample based on the distance to the decision hyperplanes with SVM. The closer to the decision boundary a sample is, the more uncertainty its classification is. In [3], Tuia et al. presented two batch-mode active learning techniques for multiclass remote sensing image classification problems. Another group chooses the representative sample which can optimize expected measures. In statistics, the problem of selecting samples to label is typically referred to as experimental design [4]. The research of optimal experimental design (OED) is concerned with the design of experiments which are expected to minimize variance of classification model. The classic optimal experimental design techniques consist of A-optimal design, D-optimal design, and E-optimal design in which they select samples to minimize the confidence region for the model parameters. Recently, Yu et al. [5] has proposed Transductive Experimental Design (TED) with both sequential and convex optimization. The TED technique selects sample to minimize the average predictive variance of the labeled function on pre-defined dataset which yielded impressive results on experiments.

However, above algorithms only consider the labeled sample, but ignore unlabeled sample. Recently, many researchers have considered the manifold learning which the sample is drawn from sampling a probability distribution that near to a sub-manifold of the ambient space. In order to exploit the underlying manifold structure, many manifold learning algorithms have been proposed, such as Locally Linear Embedding (LLE) [6], ISOMAP [7], and Laplacian Eigenmap [8]. The most important part of the manifold learning is locally invariant [9], it shows that the nearby samples have the similar labels.

In this paper, we proposed a new active learning algorithm for remote sensing image classification which benefits from recent progresses on optimal experimental design. Unlike traditional experimental design techniques, the loss function of our algorithm is defined on both labeled and unlabeled samples. It is essentially based on a graph-based learning algorithm that introduce a locality preserving into the standard least square error based loss function. The new loss function aims to

find a classifier that is locally as smooth as possible. According to the defined function, we can select the representative samples to user for labeling.

The rest of the paper is organized as follows: in Sect. 24.2, we provide a brief review of the related work. Our manifold learning algorithm is introduced in Sect. 24.3. The experimental results are presented in Sect. 24.4. Finally, we provide the concluding remarks and suggestions for future work in Sect. 24.5.

## 24.2 Related Work

### 24.2.1 Active Learning

A general active learner can be modeled as a quintuple (G, Q, S, L, and U) [10]. G is a classifier, which is trained on the labeled samples in the training set L. Q is a query function used to select the most informative samples from an unlabeled sample pool U. S is a supervisor who can assign the true class label to the selected samples from U. Initially, the training set L has few labeled samples to train the classifier G. After that, the query function Q is used to select a set of samples from the unlabeled pool U, and the supervisor S assigns a class label to each of them. Then, these new labeled samples are included into L, and the classifier G is retrained using the updated training set. The closed loop of querying and retraining continues for some predefined iterations or until a stop criterion is satisfied.

### 24.2.2 Optimal Experimental Design

Optimal Experimental Design considers learning a linear function:

$$y = \mathbf{w}^T \mathbf{x} + \varepsilon \quad (24.1)$$

where  $y$  is the observation,  $\mathbf{x}$  is the independent variable,  $\mathbf{w}$  is the weight vector, and  $\varepsilon$  is observation error ( $\varepsilon \sim N(0, \sigma^2)$ ). Suppose we have a set of labeled samples  $(\mathbf{z}_1, y_1), \dots, (\mathbf{z}_k, y_k)$ , where  $y_i$  is the label of  $\mathbf{z}_i$ . Thus, the maximum likelihood estimate for the weight vector ( $\mathbf{w}$ ) is obtained by minimizing the sum squared error

$$SSE(\mathbf{w}) = \sum_{i=1}^k (\mathbf{w}^T \mathbf{z}_i - y_i)^2 \quad (24.2)$$

Let  $Z = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k)^T$  and  $\mathbf{y} = (y_1, y_2, \dots, y_k)^T$ . The optimal solution is

$$\hat{\mathbf{w}} = (ZZ^T)^{-1}Z\mathbf{y} \quad (24.3)$$

It can be proved that  $\hat{\mathbf{w}}$  has a zero mean and a covariance matrix given by  $\sigma^2 H_{sse}^{-1}$ , where  $H_{sse}$  is the Hessian of  $SSE(\mathbf{w})$ .

$$H_{sse} = \left( \frac{\partial^2 SSE(\mathbf{w})}{\partial \mathbf{w}^2} \right) = \left( \sum_{i=1}^k \mathbf{z}_i \mathbf{z}_i^T \right) = \mathbf{Z} \mathbf{Z}^T \quad (24.4)$$

The optimal experimental design formulates the optimization problem by minimizing some measurement of estimation error derived from  $H_{sse}^{-1}$ . The three usual measures are the trace of  $H_{sse}^{-1}$ , determinant of  $H_{sse}^{-1}$ , and maximum eigenvalue of  $H_{sse}^{-1}$ . Many works on OED can be found in [5–11].

## 24.3 Laplacian Regularized Active Learning algorithm

### 24.3.1 Laplacian Regularized D-Optimal Experimental Design

Traditional experimental design makes use of only labeled samples, while neglecting the large amount of unlabeled samples. Laplacian Regularized Least Squares makes use of both labeled and unlabeled samples to explore the intrinsic geometry structure among the samples [12]. It satisfies the locally invariant idea, which assumes that if two samples  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are sufficiently close to each other, then their labels  $f(\mathbf{x}_i)$  and  $f(\mathbf{x}_j)$  are close as well [13]. Suppose there are  $k$  samples from  $m$  samples are labeled. Let  $W$  be a similarity matrix. Thus, the new loss function is defined as follows

$$RRS(\mathbf{w}) = \sum_{i=1}^k (f(\mathbf{z}_i) - y_i)^2 + \lambda \sum_{i,j=1}^m (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2 W_{ij} + \beta \|\mathbf{w}\|^2 \quad (24.5)$$

Where  $\lambda \geq 0$  and  $\beta \geq 0$  are the regularization parameters.  $y_i$  is the label of  $\mathbf{z}_i$ . In the loss function, the symmetric weights  $W_{ij}$  incurs a heavy penalty when neighboring samples are mapped far apart.

In order to model the manifold structure, we construct a nearest neighbour graph  $G$ . For each sample  $\mathbf{x}_i$ , we find its nearest neighbors and put an weight between  $\mathbf{x}_i$  and its neighbours. There are many choices of the similarity matrix. A common definition is as follows

$$W_{ij} = \begin{cases} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right) & \mathbf{x}_i \in N(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N(\mathbf{x}_i) \\ 0 & \end{cases} \quad (24.6)$$

By using the similarity matrix, the graph Laplacian is defined as  $L = D - W$  where  $D$  is a diagonal degree matrix, it is given by  $D_{ii} = \sum_j W_{ij}$ . It reflects the intrinsic geometry of the samples. Therefore, the optimal solution of (24.5) is given as follows

$$\hat{\mathbf{w}} = (\mathbf{Z}\mathbf{Z}^T + \lambda\mathbf{X}\mathbf{L}\mathbf{X}^T + \beta\mathbf{I})^{-1}\mathbf{Z}\mathbf{y} \quad (24.7)$$

Let  $S = (\mathbf{Z}\mathbf{Z}^T + \lambda\mathbf{X}\mathbf{L}\mathbf{X}^T + \beta\mathbf{I})$ , the covariance matrix of  $\hat{\mathbf{w}}$  is

$$\begin{aligned} \text{cov}(\hat{\mathbf{w}}) &= \text{cov}(S^{-1}\mathbf{Z}\mathbf{y}) \\ &= \sigma^2 S^{-1}\mathbf{Z}\mathbf{Z}^T S^{-1} \\ &= \sigma^2 S^{-1}(S - \lambda\mathbf{X}\mathbf{L}\mathbf{X}^T - \beta\mathbf{I})S^{-1} \\ &\approx \sigma^2(S^{-1} - S^{-1}(\lambda\mathbf{X}\mathbf{L}\mathbf{X}^T + \beta\mathbf{I})S^{-1}) \end{aligned} \quad (24.8)$$

where  $\mathbf{I}$  is an  $n \times n$  identity matrix.

According to previous discussion, the Laplacian regularized least square has made use of both labeled and unlabeled sample to estimate a linear fitting function that considering the intrinsic geometrical structure among the samples. A optimal design is to choose a subset  $Z \subseteq X$  which simultaneously minimizes the confidence region for  $\hat{\mathbf{w}}$  and the predictive variance of  $f(\mathbf{x})$ .

In this paper, we apply D-optimality to select the most informative samples due to its connection to the confidence region for the parameters. Since the regularization parameters are usually set to be very small, we have

$$|S^{-1} - S^{-1}(\lambda\mathbf{X}\mathbf{L}\mathbf{X}^T + \beta\mathbf{I})S^{-1}| \approx |S^{-1}| \quad (24.9)$$

where  $|\bullet|$  denote the determinant. So the smaller is, the smaller is the size of the covariance matrix.

### 24.3.2 Optimization Scheme

In the following, we describe a sequential construction of Laplacian regularized D-optimal designs. Let  $S_k$  be the hessian of  $RRS(\mathbf{w})$  after  $k$  points are selected.

$$S_k = \mathbf{Z}_k\mathbf{Z}_k^T + \lambda\mathbf{X}\mathbf{L}\mathbf{X}^T + \beta\mathbf{I} \quad (24.10)$$

where  $\mathbf{Z}_k = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k)$ . It is to show that the graph Laplacian is positive semi-definite. Therefore,  $S_k$  is positive definite and invertible. The first sample is selected such that  $|\mathbf{z}_1\mathbf{z}_1^T + \lambda\mathbf{X}\mathbf{L}\mathbf{X}^T + \beta\mathbf{I}|$  is maximized. Suppose  $k$  samples have been selected. Thus, the  $(k + 1)$ -th sample is selected such that is maximized



$$\begin{aligned}
\mathbf{z}_{k+1} &= \arg \max_{z \in X - Z_k} |S_k + \mathbf{z}\mathbf{z}^T| \\
&= \arg \max_{z \in X - Z_k} \mathbf{z}^T S_k^{-1} \mathbf{z}
\end{aligned} \tag{24.11}$$

Once the  $(k + 1)$ -th sample is selected, the inverse of  $S_{k+1}$  can be updated on the inverse of  $S_k$ , by using the Sherman-Morrison formula [14]

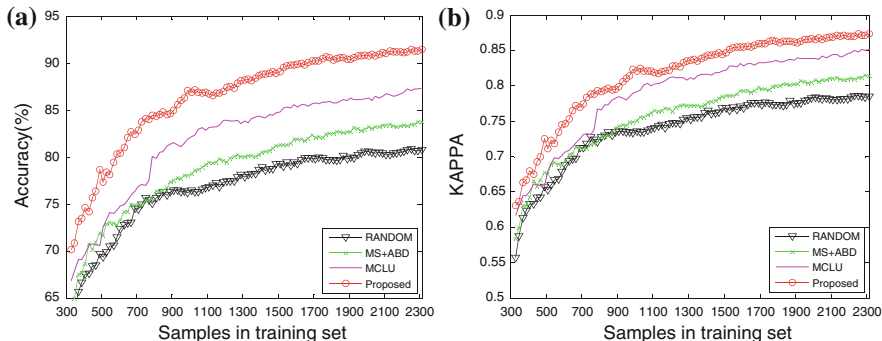
$$\begin{aligned}
S_{k+1}^{-1} &= (S_k + \mathbf{z}_{k+1}\mathbf{z}_{k+1}^T)^{-1} \\
&= S_k^{-1} - \frac{S_k^{-1}\mathbf{z}_{k+1}\mathbf{z}_{k+1}^T S_k^{-1}}{1 + \mathbf{z}_{k+1}^T S_k^{-1}\mathbf{z}_{k+1}}
\end{aligned} \tag{24.12}$$

It does not need to compute the matrix determinant and inverse. Instead, at each iteration we select a new sample such that  $\mathbf{z}^T S_k^{-1} \mathbf{z}$  is maximized and the inverse of  $S_k$  can be efficiently updated.

## 24.4 Experiments

The data set is Hyperspectral image. It is a 220-bands AVIRIS image taken over Indiana's Indian Pine test site in June 1992. The image is  $145 \times 145$  pixels, contains 16 classes representing different crops, and a total of 10,366 labeled pixels. This image is a classical benchmark to validate model accuracy and constitutes a very challenging classification problem because of the strong mixture of the class signatures. Twenty water absorption channels were removed prior to analysis. In the experiment, classes with less than 100 labeled pixels were removed, resulting thus in a 13 classes classification problem with 10,266 labeled pixels. Among the available labeled pixels, 7,000 were used for the training and candidate sets. Each experiment starts with pixels (ten per class). As for the previous image, the remaining 3,266 pixels have been used to test the generalization capabilities.

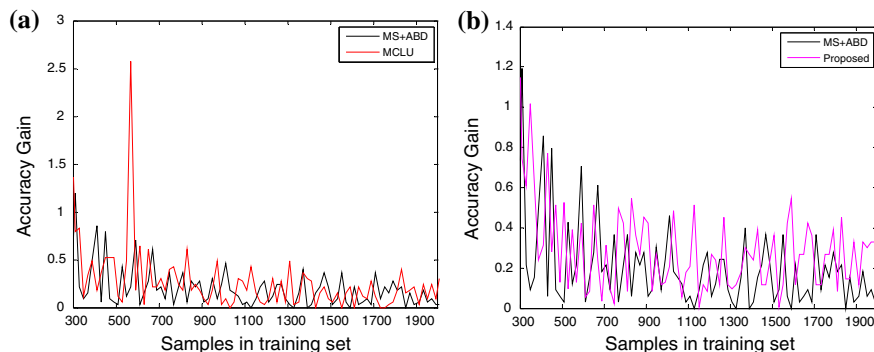
To assess the effectiveness of the proposed technique, we compared it with three other methods: (1) the random sampling (RS); (2) the margin sampling and angle based diversity (MS + ABD); (3) the multiclass-level uncertainty (MCLU). For the hyperspectral data set, initially only 130 labeled samples were included in the training set and 20 samples were selected at each iteration of active learning. The whole process was iterated until the satisfied result achieved. The active learning process runs 20 trials to reduce the random effect on the results. Figure 24.1 shows the average overall classification accuracies and error reduction provided by different methods versus the number of samples included in the training set at different iterations for the hyperspectral data set. From the figure, we can see that the proposed active learning method always resulted in higher classification accuracy than the other methods.



**Fig. 24.1** **a** The classification accuracy and **b** kappa over 20 trials runs provided by the proposed, the MCLU, the MS + ABD, and the RS methods for the hyperspectral data set

From the Fig. 24.1a, we can see that the proposed active learning technique resulted in higher classification accuracy than the other techniques. Furthermore, for hyperspectral data set, the proposed technique yielded an accuracy of 90.326 % with only 1,700 labeled samples, while using the full pool as training set which is consist of 7,000 samples, we obtained an accuracy of 91.186 %. It is worth noting that the proposed technique is effective. From the previous iterations, we can see from the figures that the improvement of the proposed method was faster than the other techniques. From the Fig. 24.1b, we can find that, in the previous step, the kappa coefficient of the RS approach that the samples were good for the classifier is higher than the MS approach. But, after some iteration, the improvement of the MS approach is better than the RS method. However, the proposed technique is always higher than the other techniques. It shows that the proposed technique is well suited to the dataset.

Figure 24.2 shows the average classification accuracy gain. From the Fig. 24.2a, the accuracy gain is similar with each other since the techniques are both based on



**Fig. 24.2** The classification accuracy gain over 20 trials runs provided by the proposed, the MS + ABD, and the MCLU methods for the hyperspectral data set

the distance to the hyperplanes. In another figure, since the training data set reached 1,450, the proposed method can efficiently reflect the intrinsic geometry of the sample space, the accuracy gain increased more than the others until convergence. It is worth noting that, even if the increase of accuracy is limited on the considered data set, we have improvements at all iterations of the active learning process.

## 24.5 Conclusion

In this paper, we have presented a novel active learning technique for solving classification problem with manifold structure. Using techniques from optimal experimental design, we select those samples such that the size of the covariance matrix of the estimated coefficients is minimized. To empirically assess the effectiveness of the proposed method, we compared it with other three active learning techniques using hyperspectral remote sensing image data set. In this comparison, we observed that the proposed method provided higher accuracy than those achieved by some of the most effective techniques presented in the literature. In this paper, we use—optimality criterion to measure the size of the covariance matrix of the coefficients. It is interesting to explore application domains where our proposed algorithm can yield good results.

**Acknowledgments** This work was supported in part by National Natural Science Foundation of China under Grant 70701013 and China Postdoctoral Science Foundation under Grant 2011M500035 and Research Fund for the Doctoral Program of Higher Education of China under Grant 20110023110002.

## References

1. Pearlman JS, Berry PS, Segal CC, Shapanski J, Beiso D, Carman SL (2003) Hyperion: a space-based imaging spectrometer. *IEEE Trans Geosci Remote Sens* 41(6):1160–1173
2. Rajan S, Ghosh J, Crawford MM (2008) An active learning approach to hyperspectral data classification. *IEEE Trans Geosci Remote Sens* 46(4):1231–1242
3. Tuia D, Ratle F, Pacifici F, Kanevski MF, Emery WJ (2009) Active learning methods for remote sensing image classification. *IEEE Trans Geosci Remote Sens* 47(7):2218–2232
4. Atkinson C, Donev AN (2007) *Optimum experimental designs with SAS*. Oxford University Press, Oxford
5. Yu K, Bi J, Tresp V (2006) Active learning via transductive experimental design. In: *Proceedings of the 23rd international conference on machine learning*, Pittsburgh
6. Roweis S, Saul L (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500):2323–2326
7. Tenenbaum J, de Silva V, Langford J (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500):2319–2323
8. Belkin M, Niyogi P (2001) Laplacian eigenmaps and spectral techniques for embedding and clustering. *Adv Neural Inf Process Syst* 14:585–591

9. Hadsell R, Chopra S, LeCun Y (2006) Dimensionality reduction by learning an invariant mapping. In: Proceedings of the 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06), pp 1735–1742
10. Settles B (2010) Active learning literature survey. CS Technical Report 1648. University of Wisconsin-Madison, Madison
11. Flaherty P, Jordan MI, Arkin AP (2005) Robust design of biological experiments. In: Advances in neural information processing systems, vol 18. Vancouver, Canada
12. Belkin M, Niyogi P, Sindhvani V (2006) Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. *J Mach Learn Res* 7:2399–2434
13. Alireza G, Hamid RR, Mohsen F, Mohammad TM, Mohammad HR (2011) Active learning from positive and unlabeled data. In: 11th IEEE international conference on data mining workshops, pp 244–250
14. Sherman J, Morrison WJ (1950) Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Ann Math Statist* 21:124–127

# Chapter 25

## Speaker Tracking Based on Audio-Visual Fusion with Unknown Noise

Jie Cao, Jun Li and Wei Li

**Abstract** In order to meet the high precision, strong robust demands of speaker tracking system, this paper proposed a new Particle filter algorithm with unknown noise statistic characteristics. The proposed algorithm estimate and correct the statistic characteristics of the unknown noise on-line by improved Sage-Husa estimator, and produce optimal distribution function with unscented Kalman filter. Finally, it realized speaker tracking problem based on audio-visual fusion in the framework of the new algorithm. Experiment results show that the method proposed in this paper has enhanced the accuracy and robustness of speaker tracking system.

**Keywords** Speaker tracking · Audio-visual fusion · Particle filter · Noise statistic characteristics

---

J. Cao  
Manufacturing Engineering Technology Research Center of Gansu,  
Lanzhou 730050, China  
e-mail: caoj@lut.cn

J. Cao  
College of Computer and Communication, Lanzhou University of Technology,  
Lanzhou 730050, China

J. Li (✉) · W. Li  
College of Electrical and Information Engineering, Lanzhou University of Technology,  
Lanzhou 730050, China  
e-mail: lljun\_lanzhou@163.com

W. Li  
e-mail: lwyz815@163.com

## 25.1 Introduction

With the rapid development of intelligent robot, teleconference and intelligent meeting scheduler system, the technique of speaker localization and tracking in intelligent environment is attracting more and more attention, which is the base of follow-up study of automatic recording, analysis and querying, as well as identity verification. The traditional methods of speaker localization and tracking includes sound source localization (SSL) based on microphone array [1] and face tracking based on computer vision [2]. But now, these single-modal approaches cannot meet the demands of high precision and strong robustness. According to the mechanism of human brain apperceiving the world, researchers study speaker tracking problem based on multi-information fusion technology and have achieved very good results.

Reference [3] describes an audio-visual 3-D person tracker that uses face detector as the visual front-end and fuses detections from multiple views to obtain the 3-D location of the person's head. If a speaker is active, the audio localization results are matched to the closest video track and continued to be tracked. If there is no match with the video tracks, the audio track is tracked separately. The results indicate that though the video face detection yields consistent results, the fusion of audio localization information does not perform well. In fact, with the addition of audio information the results are worse than the video-only results.

Reference [4] presents a probabilistic method for audio-visual (AV) speaker tracking, using an uncalibrated wide-angle camera and a microphone array. The algorithm fuses 2-D object shape and audio information via importance particle filters, allowing for the asymmetrical integration of AV information in a way that efficiently exploits the complementary features of each modality.

Reference [5] presents an interesting particle filtering framework which incorporates the audio and visual detections into the particle filtering framework. However, the tracking framework presented in [5] does not correspond to a 3-D tracker. The camera views are stitched to obtain a panoramic view of the room in which subjects are tracked. An advantage of this system is that the cameras need not be accurately calibrated. However, this setup places restriction on the positions that the subjects can occupy and is difficult to generalize to new scenes especially when a large number of people participate in meetings and lectures.

Although these study obtained very good results, but all of them were under the hypothesis of noise statistical characteristics was known accurately. In most situations, accurate system noise statistic characteristics cannot obtain in advance, which is a bottleneck technology. In order to overcome this difficulty, this paper proposed an adaptive Particle Filtering algorithm. The main idea of the proposed method in this paper is that use improved Sage-Husa suboptimal unbiased maximum posteriori (MAP) recursively estimate system noise covariance. Combined with UKF method, system noise statistic characteristics can be estimated and corrected online. Thus, the proposal distribution function containing latest observing information is obtained, which improve the accuracy of speaker motion model and tracking accuracy of system, as well as the anti-jamming ability of

system. At last, speaker tracking based on audio-visual information fusion is achieved under this new algorithm framework.

## 25.2 The Principle Algorithm

### 25.2.1 Standard Particle Filter Algorithm

The particle filter (PF) is an alternative Bayesian approach that performs sequential Monte Carlo (SMC) estimation. It supports nonlinear and non-Gaussian state-space models and hence is suitable for solving the localization problem. In this approach, no particular model is assigned to the posterior density function (PDF). PF represents the required PDF by a set of random sampled particles with associated weights. Normally, the thought of standard particle filter is that using a series of random sample  $\chi_k^{(i)}$  and the corresponding weights  $\omega_k^{(i)}$  to represent the posterior probability density or filtering probability density of the speaker's position [6]:

$$p(\chi_k|Y_k) = \sum_{i=1}^N \omega_k^{(i)} \delta(\chi_k - \chi_k^{(i)}) \quad (25.1)$$

In order to calculate conveniently, standard particle filter algorithm uses state transfer function  $p(\chi_k|\chi_{k-1})$  as the important probability density function. Combining with the system state equation to generate random sampling particles  $\{\chi_k^{(i)}\}_{i=1}^N$ , and uses the beam energy output  $p(Y_k|\chi_k^{(i)})$  as the likelihood function. The particle weights can be calculated as:

$$\omega_k^{(i)} \propto \omega_{k-1}^{(i)} \frac{p(Y_k|\chi_k^{(i)})p(\chi_k^{(i)}|\chi_{k-1})}{q(\chi_k^{(i)}|\chi_{1:k-1}, Y_k)} \quad (25.2)$$

$$= \omega_{k-1}^{(i)} p(\chi_k^{(i)}) \quad (25.3)$$

The minimum variance estimation of speaker's position  $\chi_k$  can be expressed as:

$$\hat{\chi}_k = \sum_{i=1}^N \chi_k^{(i)} \omega_k^{(i)} \quad (25.4)$$

Although the method that using the state transfer function as important sampling function is easy to carry out, but it did not add the latest measurement information into the proposal distribution function. It is easy to cause matching error of the system model after multiple iterative. And finally, particle weights come to degenerate, and the precision of filter estimation will reduce and even disable [7].

### 25.2.2 UKF Algorithm

In order to resolve nonlinear filtering problem, Refs. [8, 9] proposed an unscented Kalman filter (UKF) method based on the unscented transformation. The specific algorithm just as follows:

- (1) For a given  $\hat{x}_{k-1|k-1}$  and  $\mathbf{P}_{k-1|k-1}$ , using the method of U transformation to acquire the value of one state step prediction  $\hat{x}_{k|k-1}$  and the covariance matrix  $\mathbf{P}_{k|k-1}$  of forecast error.

① Calculate the point of Sigma:  $\chi_{k-1|k-1}^i$ ,  $i = 0, 1, 2, \dots, 2n$

$$\begin{cases} \chi_{k-1|k-1}^0 = \hat{x}_{k-1|k-1} \\ \chi_{k-1|k-1}^i = \hat{x}_{k-1|k-1} + \left( \sqrt{(n+\lambda)\mathbf{P}_{k-1|k-1}} \right)_i, & i = 1, 2, \dots, n \\ \chi_{k-1|k-1}^i = \hat{x}_{k-1|k-1} - \left( \sqrt{(n+k)\mathbf{P}_{k-1|k-1}} \right)_i, & i = n+1, n+2, \dots, 2n \end{cases} \quad (25.5)$$

② Calculating Sigma point  $\chi_k^i$ ,  $i = 0, 1, 2, \dots, 2n$ , through the spread of the state evolution equation, namely:

$$\begin{cases} \chi_k^i = f\left(\chi_{k-1|k-1}^i\right) \\ \hat{x}_{k|k-1} = \sum_{i=0}^{2n} W_i^m \chi_k^i \\ \mathbf{P}_{k|k-1} = \sum_{i=0}^{2n} W_i^c (\chi_k^i - \hat{x}_{k|k-1})(\chi_k^i - \hat{x}_{k|k-1})^T + \mathbf{Q}_{k-1} \end{cases} \quad (25.6)$$

- (2) Computing Sigma points by the spread of the measuring information.

① Computing Sigma points  $\hat{x}_{k|k-1}$  and  $\mathbf{P}_{k|k-1}$  through the spread of the measurement equation as follows:

$$\begin{cases} \chi_k^0 = \hat{x}_{k-1|k-1} \\ \chi_k^i = \hat{x}_{k|k-1} + \left( \sqrt{(n+\lambda)\mathbf{P}_{k-1|k-1}} \right)_i, & i = 1, 2, \dots, n \\ \chi_k^i = \hat{x}_{k|k-1} - \left( \sqrt{(n+k)\mathbf{P}_{k-1|k-1}} \right)_i, & i = n+1, n+2, \dots, 2n \end{cases} \quad (25.7)$$



② Calculate the output of the one step prediction value:

$$\begin{cases} y_k^i = h(\chi_k^i), & i = 0, 1, 2, \dots, 2n \\ \hat{y}_{k|k-1} = \sum_{i=0}^{2n} W_i^m y_k^i \\ \mathbf{P}_y = \sum_{i=0}^{2n} W_i^c (y_k^i - \hat{y}_{k|k-1}) (y_k^i - \hat{y}_{k|k-1})^T + \mathbf{R}_k \\ \mathbf{P}_{xy} = \sum_{i=0}^{2n} W_i^c (\chi_k^i - \hat{x}_{k|k-1}) (\chi_k^i - \hat{x}_{k|k-1})^T \end{cases} \quad (25.8)$$

(3) After getting the new measurement information, then filtering the updating as follows:

$$\begin{cases} \hat{x}_{k|k} = \hat{x}_{k|k-1} + \mathbf{K}_k (y_k - \hat{y}_{k|k-1}) \\ \mathbf{K}_k = \mathbf{P}_{xy} \mathbf{P}_y^{-1} \\ \mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{P}_y^{-1} \mathbf{K}_k^T \end{cases} \quad (25.9)$$

where  $\mathbf{K}_k$  stands for filtering gain matrix.

### 25.2.3 Statistic Characteristics Estimation of System Noise

According to the description of literature [10], we can get the estimation of the statistical characteristics of system noise based on Sage-Husa estimator. Just as follows:

$$\hat{q}_k = \frac{1}{k} \sum_{j=1}^k [\hat{x}_{j|k} - \Phi_{j-1} \hat{x}_{j-1|k}] \quad (25.10)$$

$$\hat{Q}_k = \frac{1}{k} \sum_{j=1}^k [\hat{x}_{j|k} - \Phi_{j-1} \hat{x}_{j-1|k} - \hat{q}_k] \times [\hat{x}_{j|k} - \Phi_{j-1} \hat{x}_{j-1|k} - \hat{q}_k]^T \quad (25.11)$$

The effective method to overcome the weights degradation problem of particle filtering algorithm is that to enhance the correctional ability of the latest measurements. Therefore, in the noise estimating stage, this paper assigned different weights to the latest measurement information and past information. Just as follows:

$$\hat{q}_k = (1 - d_{k-1}) \hat{q}_{k-1} + d_{k-1} [\hat{x}_{k|k} - \Phi_k \hat{x}_{k-1|k-1}] \quad (25.12)$$

$$\hat{Q}_k = (1 - d_{k-1}) \hat{Q}_{k-1} + d_{k-1} [\mathbf{K}_k \mathbf{v}_k \mathbf{v}_k^T \mathbf{K}_k^T + \mathbf{P}_{k|k} - \Phi_k \mathbf{P}_{k-1|k-1} \Phi_k^T] \quad (25.13)$$

where  $d_{k-1} = (1 - b)/(1 - b^k)$ ,  $j = 0, 1, \dots, k - 1$ .  $b$  is called as the forgetting factor, according to experience value in the related research, usually the range of the value of  $b$  is  $0.95 < b < 0.99$ .

## 25.3 Speaker Tracking Based on Audio-Visual Fusion

### 25.3.1 Motion Modeling

Speaker representations and state-spaces defined either on the image plane or in 3-D space are sensible choices. In our case, we have chosen a subspace of transformations comprising translation  $T^x, T^y$ . Furthermore, a second-order auto-regressive dynamical model is defined by  $\mathbf{x}_t = (x_t, x_{t-1})^T$ , and  $x_t = (T_t^x, T_t^y)$ , the dynamical model is defined by  $\mathbf{x}_t = A\mathbf{x}_{t-1} + Bw_t$ , where  $A, B$  are the parameters of the model, and  $w$  is process noise with unknown statistical characteristics.

### 25.3.2 Visual Observations Model

The observation model assumes that shapes are embedded in clutter. Edge-based measurements are computed along  $L$  normal lines to a hypothesized contour, resulting in a vector of candidate positions for each line,  $y_t^l = \{v_m^l\}$  relative to the point lying on the contour  $v_0^l$ . With some usual assumptions, the observation likelihood for  $L$  normal lines can be expressed as

$$p(y_t^{vid}|x_t) \propto \prod_{l=1}^L p(y_t^l|x_t) \propto \prod_{l=1}^L \max\left(K, \exp\left(-\frac{\|\hat{v}_m^l - v_0^l\|}{2\sigma^2}\right)\right) \quad (25.14)$$

where  $\hat{v}_m^l$  is the nearest edge detected on the  $l$ th line, and  $K$  is a constant introduced when no edges are detected.

### 25.3.3 Audio Observation Model

In general, audio localization methods rely on estimation of the delay between the time of arrival of a signal on a pair of microphones. We define the vector of theoretical time delays associated with a 3-D location  $\mathbf{x}$  as  $\tau^{1:M,\mathbf{x}} = \{\tau^{m,\mathbf{x}}\} \cong \tau^{\mathbf{x}}$ , where  $\tau^{m,\mathbf{x}}$  is the delay (in samples) between the microphone in pair  $m$ ,

$$\tau^{m,\mathbf{x}} = \frac{(\|\mathbf{x} - M_1^m\| - \|\mathbf{x} - M_2^m\|)f_s}{c} \quad (25.15)$$

where  $M_1^m$  and  $M_2^m$  are the locations of microphones in pair  $m$ , and  $f_s$  is the sampling frequency. In practice, each time delay estimate  $\hat{\tau}_t^m$  is calculated from the generalized cross-correlation (GCC) [11]. A phase transform (PHAT) is applied to improve the robustness to reverberation, and the GCC is interpolated to achieve sub-sample precision (details in [12]). Then, given an vector  $\hat{\tau}_t \cong \{\hat{\tau}_t^m\}$  of

observed time delay estimates, the distribution of the observation given a speaker at location  $\mathbf{x}$  can be modeled as  $p(\hat{\tau}_t|\tau^x) = N(\tau^x, \Sigma^x)$ , where  $\Sigma^x$  is the covariance matrix, chosen to be independent of location. The location estimate can then be defined according to the maximum likelihood (ML) criterion as  $\hat{\mathbf{x}} = \arg \max p(\hat{\tau}_t|\tau^x)$ . The location estimate for each frame is found by a dynamic search over  $p(\hat{\tau}_t|\tau^x)$ . The localization estimate for each frame is found by a dynamic search over  $p(\hat{\tau}_t|\tau^x)$  through a uniform grid of room locations. To eliminate low confidence values, estimates whose likelihood falls below that of a uniform distribution are labeled as silence, meaning the audio observations contain discontinuities. To synchronize frame rates, multiple audio frames are merged by selecting only the ML location across frames.

### 25.3.4 Audio-Visual Fusion

Assuming independence, we define  $i_t(x_t) = i_t(T_t^x, T_t^y) = N(u_t, \Sigma_t)$ . The mean  $u_t = (u_t^x, u_t^y)$  consists of the project 3-D audio estimate onto the image plane  $C(\hat{\mathbf{x}}_t)$  and the unit scale. The covariance matrix  $\Sigma_t$  is diagonal, with translation components proportional to the mean head size in the training set, and with scaling component equal to the variance in scale of head sizes. In case of silence, no IS function exists, so the filter draws samples only from the dynamical model. The importance function is also used for the audio-based observation likelihood,

$$p(y_t^{aud}|x_t) \propto i_t(x_t) \quad (25.16)$$

in case there is audio, and it is a fixed constant otherwise. Then observations are combined in a standard approach,

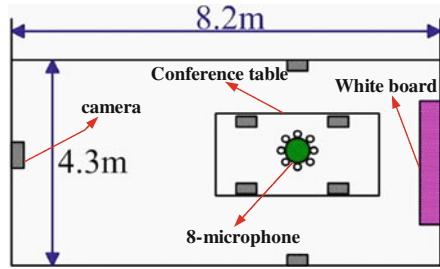
$$p(y_t|x_t) = p(y_t^{vid}|x_t)p(y_t^{aud}|x_t) \quad (25.17)$$

## 25.4 Experimental Analysis

### 25.4.1 Experimental Setup

Audio-visual recording were made in a meeting room with one wide-angle camera on a wall and an 8-microphone array on the table (Fig. 25.1). Video was captured at 25 fps, while audio was recorded at 16 kHz, with features estimated at 62.5 fps. Images were processed in CIF format, so a human head is about  $20 \times 35$  pixels (1 pixel  $\approx 8$  mm). Parameters for the visual tracker (dynamics and observations) have been kept fixed for all experiments.

Fig. 25.1 Meeting room



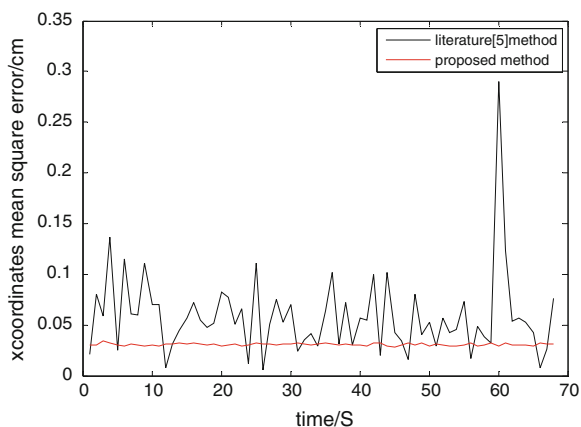
### 25.4.2 Results Analysis

In order to give a comprehensive evaluation of the proposed method, here we carry out three experiments separately: single speaker tracking based on simple background; speaker tracking under visual clutter environment in meeting experiment. The tracking results were compared comprehensively between proposed method and literature [5] method. The tracking results just as Figs. 25.2, 25.3, 25.4, 25.5, 25.6 and 25.7.

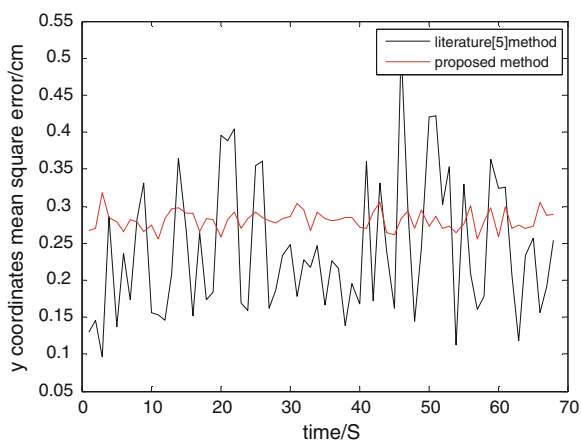


Fig. 25.2 Scene 1: Tracking results in simple background

**Fig. 25.3** RMSE of x coordinate of scene 1



**Fig. 25.4** RMSE of y coordinate of scene 1

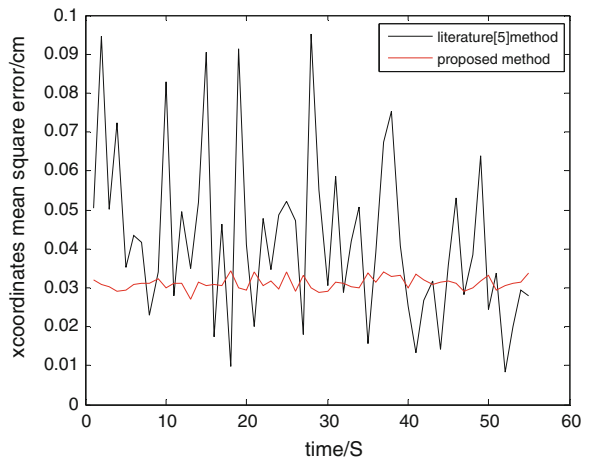


The tracking results indicate that the proposed method and literature [5] method both maintaining a good tracking precision in simple experiment scene, we can get these from Figs. 25.3 and 25.6. But Figs. 25.4, 25.5 and 25.7 show that the proposed method maintained a good tracking stability, because the proposed method in this paper estimate the noise statistical characteristics online, and the influence of environmental noise disturbance is small.

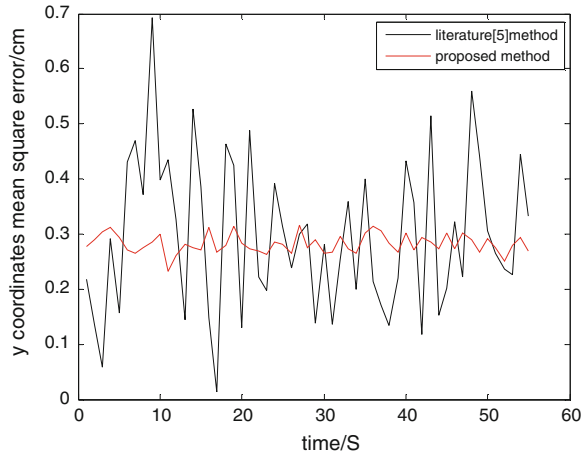


Fig. 25.5 Scene 2: Tracking results under visual clutter

Fig. 25.6 RMSE of x coordinate of scene 2



**Fig. 25.7** RMSE of y coordinate of scene 2



## 25.5 Conclusion

According to the speaker tracking problem, this paper proposed a kind of nonlinear filtering method with unknown statistical properties. Through the improved Sage-Husa estimator to estimate the system noise statistical properties online, and fuse the estimated information into the propose distribution function by the UKF method, effectively improve the system's overall filtering precision. At last, this paper realized speaker tracking problem based on audio and visual fusion in the new algorithm framework, and achieved good tracking performance.

**Acknowledgments** The research was supported by Nation Natural Science Foundation of China (61263031), Natural Science Foundation of Gansu province of China (1010RJA046).

## References

1. Cobos M, Lopez JJ, Martinez D (2011) Two-microphone multi-speaker localization based on a Laplacian mixture model. *Digital Signal Proc* 21(1):66–76
2. Shivappa ST, Trivedi M, Rao D (2011) Audio-visual Information fusion in human computer interfaces and intelligent environments: a survey. *IEEE Proc* 98(10):1680–1691
3. Shivappa ST, Rao BD, Trivedi MM (2010) Audio-visual fusion and tracking with multilevel iterative decoding: framework and experimental evaluation. *IEEE J Sel Top Signal Proc* 4(5):882–894
4. Gatica-Perez D, Lathoud G, McCowan I, Odobez J-M, Moore D (2003) Image processing. In: *Proceedings of ICIP 2003 international conference*, vol 3, issue 2, pp 5–8
5. Perez DG, Lathoud G, Odobez JM, Cowan IM (2007) Audio-visual probabilistic tracking of multiple speakers in meetings. *IEEE Trans Audio Speech Lang Process* 15(2):601–615
6. Checka N, Wilson KW, Siracusa MR et al (2004) Multiple person and speaker activity tracking with a particle filter. In: *Proceedings of IEEE international conference on acoustics, speech, and signal processing*, May 2004

7. Gordon NJ, Salmond DJ, Smith AFM (1993) Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEEE Proc Radar Signal Process* 140(2):107–113
8. Julier SJ, Uhlmann JK (2004) Unscented filtering and nonlinear estimation. *Proc IEEE* 92(3):401–422
9. Wan EA, Merwe R (2000) The unscented Kalman filter for nonlinear estimation. In: *Proceedings of the international symposium on adaptive systems for signal processing, communications and control*, Alberta, Canada, pp 153–158
10. Shi Y, Han C-Z (2011) Adaptive UKF method with applications to target tracking. *Acta Automatica Sin* 37(6):755–759 (in Chinese)
11. Cao J, Zheng J (2012) Speaker tracking based on audio-video information fusion. *Comput Eng Appl* 48(13):118–124 (in Chinese)
12. Blauth MV, Claudio PJ et al (2012) Voice activity detection and speaker localization using audiovisual cues. *Pattern Recogn Lett* 33(4):373–380



# Chapter 26

## The Analysis of Epidemic Disease Propagation in Competition Environment

Mingsheng Hu, Suimin Jia, Qiaoling Chen, Zhijuan Jia and Liu Hong

**Abstract** An autonomous SIRS epidemic model of two competitive species is established in this paper, in which One kind of disease can survive and have the chance of cross-infection, not only the disease of the cross-infection but also the additional disease death rates are considered in this model, through the analysis of this kind of model, we can gain the threshold value condition of the stability of equilibrium. Even more it can see the global stability of the model through simulation number value.

**Keywords** SIRS · Local asymptotic stability · Global asymptotic stability · Numerical simulation

### 26.1 Introduction

The epidemic disaster has historically been the enemy of human health and the different harmful biological viruses is huge threat to human health.

For nearly 20 years, the research progress of epidemic disaster prevention is quickly in the international, many mathematical model is used to the analysis variety of infectious diseases. Much of the these mathematical model apply to study the general laws of epidemic's. It is based on the characteristics of population growth, the occurrence of the disease, development of law, and the relevant social factors such as infectious diseases, established the mathematical models to reflect the dynamic characteristics. By qualitative and quantitative analysis and numerical simulation of the model to display the process of the disease, and reveal the epidemic law, it predicts change and development trends, analyze the causes of disease epidemic and key factors, seeking its prevention and control of the optimal

---

M. Hu (✉) · S. Jia · Q. Chen · Z. Jia · L. Hong  
College of Information Science and Technology, Zhengzhou Normal University, No.6,  
Yincai Street, Huiji District, Zhengzhou City, Henan, China  
e-mail: hero\_jack@163.com

strategy which provide a theoretical basis and quantitative basis for people to formulate a prevention strategy, but also some investigate about such as plague, measles, malaria, tuberculosis, AIDS, and many other specific diseases. Currently theoretical analysis of the mathematical model based on epidemic disease, has been more extensive studies [1–10], but as an important means of model testing and prediction. Computer simulation not only improve hazard early warning and emergency decision support level but also enhance the overall prevention and control of major urban infectious diseases and other emergencies which play an important role, it can find the simple solution put the complex abstract theoretical into easily understand numerical sequence or graphical presentate and make the number of huge parameter filtering and dynamic process simulation into possible.

A species in nature can not exist in isolation. it always exist interaction with other due to competition for food, resources and space, so to consider the interaction of two or more populations is very important in the infectious disease model, most of the work is for prey system [1–4], as for the competitive system, some scholars have also studies [1, 5, 6]. An autonomous SIRS epidemic model of two competitive species is established in this paper, in which One kind of disease can exist and have the chance of cross-infection, not only the disease of the cross-infection but also the additional disease death rates are considered in this model, through the analysis of this kind of model, we can gain whenever the stability of equilibrium of the threshold value. And through numerical simulation, more intuitive to see the model of global stability.

## 26.2 Model Establishment

Let us consider the following SIRS infectious disease model

$$\left\{ \begin{array}{l} \dot{N}_1 = r_1 \left( 1 - \frac{N_1}{K_1} \right) N_1 - aN_1N_2 \\ \dot{S}_1 = \left( b_1 - \frac{a_1N_1r_1}{K_1} \right) N_1 - \left[ d_1 + (1 - a_1) \frac{N_1r_1}{K_1} \right] S_1 - aN_2S_1 - S_1(\beta_{11}I_1 + \beta_{12}I_2) + \delta_1R_1 \\ \dot{I}_1 = S_1(\beta_{11}I_1 + \beta_{12}I_2) - \gamma_1I_1 - \left[ d_1 + (1 - a_1) \frac{N_1r_1}{K_1} \right] I_1 - aN_2I_1 \\ \dot{R}_1 = \gamma_1I_1 - \left[ d_1 + (1 - a_1) \frac{N_1r_1}{K_1} \right] R_1 - \delta_1R_1 - aN_2R_1 \\ \dot{N}_2 = kaN_1N_2 - d_2N_2 \\ \dot{S}_2 = kaN_1N_2 - S_2(\beta_{21}I_1 + \beta_{22}I_2) + \delta_2R_2 - d_2S_2 - nN_1S_2 \\ \dot{I}_2 = S_2(\beta_{21}I_1 + \beta_{22}I_2) - d_2I_2 - \gamma_2I_2 - nN_1I_2 \\ \dot{R}_2 = \gamma_2I_2 - d_2R_2 - \delta_2R_2 - nN_1R_2 \end{array} \right. \quad (26.1)$$

Here  $\left(b_1 - \frac{a_1 N_1 r_1}{K_1}\right)$ ,  $\left[d_1 + (1 - a_1) \frac{N_1 r_1}{K_1}\right]$  is the births and deaths of the prey. The total predation rate is  $aN_1N_2$ , and the conversion rate is denoted by  $k$ ,  $\beta_{ii}$  is intraspecific infection rate,  $\gamma_i$  is the removal rate.  $n$  is interspecific competition factor, and  $\delta_i$  is the rate constant for the recovery becoming susceptible and  $d_i$  is the death rate of the predator.  $\beta_{ij}$  is Interspecific infection rate.

We put  $N_i = S_i + I_i + R_i (i = 1, 2)$ , in this model (26.1) that leads to the following set of differential equations:

$$\begin{cases} \dot{N}_1 = \left[ r_1 \left( 1 - \frac{N_1}{K_1} \right) - aN_2 \right] N_1 \\ \dot{I}_1 = (\beta_{11}I_1 + \beta_{12}I_2)(N_1 - I_1 - R_1) - \gamma_1 I_1 - \left[ d_1 + (1 - a_1) \frac{r_1 N_1}{K_1} \right] I_1 - aN_2 I_1 \\ \dot{R}_1 = \gamma_1 I_1 - \left[ d_1 + (1 - a_1) \frac{r_1 N_1}{K_1} \right] R_1 - aN_2 R_1 - \delta_1 R_1 \\ \dot{N}_2 = (k a N_1 - d_2) N_2 \\ \dot{I}_2 = (\beta_{21}I_1 + \beta_{22}I_2)(N_2 - I_2 - R_2) - d_2 I_2 - \gamma_2 I_2 - n N_2 I_2 \\ \dot{R}_2 = \gamma_2 I_2 - d_2 R_2 - \delta_2 R_2 - n N_1 R_2 \end{cases} \quad (26.2)$$

By examining (26.2) on each boundary surface, we can show that the region  $D = \{(I_1, R_1, N_1, I_2, R_2, N_2)^T | 0 \leq I_i, 0 \leq R_i, I_1 + R_1 \leq N_1 \leq K_1, I_2 + R_2 \leq N_2\}$  is positively invariant with respect to the model. System (26.2) is not have analytical solution. We make use of MATLAB obtain the system (26.2)'s numerical solution.

From Fig. 26.1 we can get conclusion of that: with the growth of the time  $N_1, N_2, I_1, I_2, R_1, R_2$  will tend to be stable.

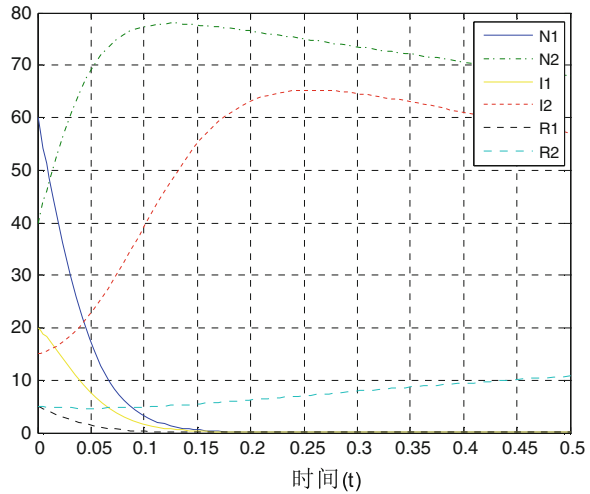
## 26.3 Equilibrium Points and Their Stability

Let us consider a basic predator-prey model

$$\begin{aligned} \dot{N}_1 &= r_1 \left( 1 - \frac{N_1}{K_1} \right) N_1 - aN_1N_2 = \left[ r_1 \left( 1 - \frac{N_1}{K_1} \right) - aN_2 \right] N_1 \\ \dot{N}_2 &= k a N_1 N_2 - d_2 N_2 = (k a N_1 - d_2) N_2 \end{aligned} \quad (26.3)$$

which is the modification of the classic Lotka-Volterra equations with density-dependent logistic growth of the prey. This model was formulated and explained by Pielou [7]. The total predation rate is  $aN_1N_2$ , and the feeding efficiency in turning predation into new predators is  $k$ ,  $(N_{1E}, N_{2E})$  is the non-negative equilibria of the system.

**Fig. 26.1** Numerical simulation of the epidemic model in competition environment



**Theorem 1** In the system of (26.3):

if  $N_1(0) > 0$  and  $d_2/(kaK_1) < 1$ , or  $N_1(0) > 0$  and  $N_2(0) = 0$ , then  $N_1 \rightarrow K_1$ ,  $N_2 \rightarrow 0$ , as  $t \rightarrow \infty$ , if  $N_1(0) > 0$ ,  $N_2(0) > 0$ , and  $d_2/kaK_1 < 1$ . Then  $N_1 \rightarrow N_{1E} = d_2/ka$ ,  $N_2 \rightarrow N_{2E} = r_1(1 - d_2/kaK_1)/a$  as  $t \rightarrow \infty$ .

The following results are used later in this paper.

Consider the systems:

$$\dot{x} = f(t, x) \tag{26.4}$$

$$\dot{y} = g(y) \tag{26.5}$$

where  $f$  and  $g$  are continuous and locally Lipschitz in  $x$  in  $R^n$  and solutions exist for all positive time. Eq. (26.4) is called asymptotically autonomous with limit Eq. (26.5) if  $f(x) \rightarrow g(x)$  as  $t \rightarrow \infty$  uniformly for  $x$  in  $R^n$ :

**Lemma 1** [8] Let  $e$  be a locally asymptotically stable equilibrium of (26.5) and  $\omega$  be the  $\omega$ -limit set of a forward bounded solution  $x(t)$  of (26.4). If  $\omega$  contains a point  $y_0$  such that the solution of (26.5) with  $y(0) = y_0$  converges to  $e$  as  $t \rightarrow \infty$ ; then  $\omega = \{e\}$ ; i.e.  $x(t) \rightarrow e$ , as  $t \rightarrow \infty$ .

**Corollary 1** If solutions of system (26.4) are bounded and the equilibrium  $e$  of the limit system (26.5) is globally asymptotically stable, then any solution  $x(t)$  of system (26.4) satisfies  $x(t) \rightarrow e$  as  $t \rightarrow \infty$ .

### 26.3.1 Equilibrium Points and Their Stability

The non-negative equilibria of the system (26.2) are :

$$\begin{aligned} P_0 &= (0, 0, 0, 0, 0, 0), P_1 = (0, 0, K_1, 0, 0, 0) \\ P_2 &= \left( \frac{K_1(b_1 - r_1 a_1)(1 - \frac{1}{R_0})}{\gamma_1 + b_1 - r_1 a_1}, \frac{K_1 \gamma_1 (1 - \frac{1}{R_0})}{\gamma_1 + b_1 - r_1 a_1}, K_1, 0, 0, 0 \right) \\ P_3 &= (0, 0, N_{1E}, 0, 0, N_{2E}), P_4 = (0, 0, N_{1E}, I_{2E}, R_{2E}, N_{2E}). \end{aligned}$$

where

$$\begin{aligned} R_0 &= \frac{\beta_{11} K_1}{\gamma_1 + b_1 - r_1 a_1}, N_{1E} = \frac{d_2}{ka}, N_{2E} = \frac{r_1(1 - \frac{d_2}{kaK_1})}{a}, \\ I_{2E} &= \frac{N_{2E}(d_2 + \delta_2 + nN_{1E})(1 - \frac{1}{R_2})}{\gamma_2 + d_2 + \delta_2 + nN_{1E}}, R_{2E} = \frac{\gamma_2 N_{2E}(1 - \frac{1}{R_2})}{\gamma_2 + d_2 + \delta_2 + nN_{1E}}, \\ R_1 &= \frac{\beta_{11} N_{1E}}{\gamma_1 + d_1 + (1 - a_1)r_1 d_2 / kaK + aN_{2E}} \end{aligned}$$

**Theorem 2** The system (26.2) has the following conclusion:

- (1)  $P_0$  is globally asymptotically stable in  $\Omega_1$ .
- (2) if  $\frac{d_2}{kaK_1} \geq 1, R_0 \leq 1, P_1$  is globally asymptotically stable in  $\Omega_2$ .
- (3) if  $\frac{d_2}{kaK_1} \geq 1, R_0 > 1, P_2$  is globally asymptotically stable in  $\Omega_3$ .
- (4) if  $\frac{d_2}{kaK_1} < 1, R_1 \leq 1, R_2 \leq 1, P_3$  is globally asymptotically stable in  $\Omega_4$ .
- (5) if  $\frac{d_2}{kaK_1} < 1, R_1 \leq 1, R_2 > 1, P_4$  is globally asymptotically stable in  $\Omega_5$ .

Where:

$$\begin{aligned} \Omega_1 &= \{(I_1, R_1, N_1, I_2, R_2, N_2)^T | 0 \leq I_i, 0 \leq R_i, I_1 + R_1 \leq N_1 \leq K_1, I_2 + R_2 \leq N_2, N_1(0) = 0\} \\ \Omega_2 &= \{(I_1, R_1, N_1, I_2, R_2, N_2)^T | 0 \leq I_i, 0 \leq R_i, I_1 + R_1 \leq N_1 \leq K_1, I_2 + R_2 \leq N_2, N_1(0) > 0\} \\ \Omega_3 &= \{(I_1, R_1, N_1, I_2, R_2, N_2)^T | 0 \leq I_i, 0 \leq R_i, I_1 + R_1 \leq N_1 \leq K_1, I_2 + R_2 \leq N_2, I_1(0) > 0\} \\ \Omega_4 &= \{(I_1, R_1, N_1, I_2, R_2, N_2)^T | 0 \leq I_i, 0 \leq R_i, I_1 + R_1 \leq N_1 \leq K_1, I_2 + R_2 \leq N_2, N_1(0) > 0, N_2(0) > 0\} \\ \Omega_5 &= \{(I_1, R_1, N_1, I_2, R_2, N_2)^T | 0 \leq I_i, 0 \leq R_i, I_1 + R_1 \leq N_1 \leq K_1, I_2 + R_2 \leq N_2, N_1(0) > 0, I_2(0) > 0\} \end{aligned}$$

Proof:

- (1) The conclusion is clearly established, Hence the proof.
- (2) At  $P_1$  the eigenvalue of the Jacobian matrix is

$$\begin{aligned} \lambda_1 &= \beta_{11}K - \gamma_1 - (b_1 - r_1 a_1), & \lambda_2 &= -(b_1 - r_1 a_1), & \lambda_3 &= -r \\ \lambda_4 &= -d_2 - \gamma_2, & \lambda_5 &= -d_2 - \delta_2, & \lambda_6 &= -d \end{aligned}$$

These 6 eigenvalues are negative when  $\frac{d_2}{kaK_1} \geq 1, R_0 \leq 1$ , so  $P_1$  is locally.

Asymptotically, let's proof  $P_1$  is globally asymptotically stable. When  $\frac{d_2}{kaK_1} \geq 1$ , by the Theorem 1  $N_1 \rightarrow K, N_2 \rightarrow 0$  if  $N_1(0) > 0$  and  $N_2 \rightarrow 0, I_2 \rightarrow 0, R_2 \rightarrow 0$  as  $t \rightarrow \infty$ , so Progressive autonomous system of (26.2):

$$\begin{aligned} \dot{I}_1 &= \beta_{11}(K_1 - I_1 - R_1)I_1 - (\gamma_1 + b_1 - r_1a_1)I_1 \\ \dot{R}_1 &= \gamma_1I_1 - (b_1 - r_1a_1)R_1 \end{aligned}$$

From corollary 1, we can obtain that the bounded solution of (26.2) will approach to global asymptotic stability of the (26.3)'s equilibrium.  $R_0 \leq 1$ , from corollary 1  $I_1 \rightarrow 0, R_1 \rightarrow 0$  as  $t \rightarrow \infty$  we can see  $P_1$  is global attractive. so  $P_1$  is globally asymptotically stable.

(3) At  $P_2$  the Jacobian matrix of the system (26.2) is

$$\begin{aligned} J(P_2) &= r_1d_2c_{11}(\gamma_1 + b_1 - r_1a_1)[(d_2 + \delta_2 + nK_1)(d_2 + \gamma_2) + \beta_{21}I_1(\gamma_2 + \delta_2 \\ &\quad + d_2 + nK_1)] + r_1d_2K_1 \frac{1}{R_0} \beta_{21}I_1(b_1 - r_1a_1)(d_2 + \delta_2 + nK_1) \end{aligned}$$

where  $c_{11} = (b_1 - r_1a_1)(1 - R_0)$   $c_{22} = r_1(\frac{1}{R_0} - 1)(1 - a_1) / \gamma_1 + b_1 - r_1a_1$

We get  $trJ(P_2) < 0, \det J(P_2) > 0$  so  $P_2$  is locally asymptotically, let's proof  $P_2$  is globally asymptotically stable. By reason that  $N_1(0) > 0, \frac{d_2}{kaK_1} \geq 1$ , from corollary 1  $N_1 \rightarrow K_1, N_2 \rightarrow 0$  as  $t \rightarrow \infty$ . Obviously, we can get that  $I_2 \rightarrow 0, R_2 \rightarrow 0$ , similarly available its autonomous limit system

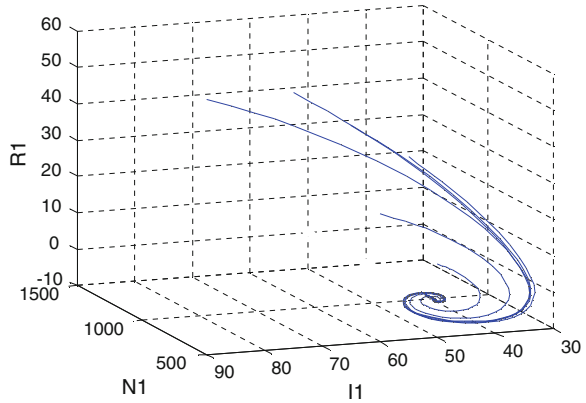
$$\begin{aligned} \dot{I}_1 &= \beta_{11}(K_1 - I_1 - R_1)I_1 - (\gamma_1 + b_1 - r_1a_1)I_1 \\ \dot{R}_1 &= \gamma_1I_1 - (b_1 - r_1a_1)R_1 \end{aligned}$$

By  $I_1(0) \neq 0$ , we get  $I_1 \rightarrow \frac{K_1(b_1 - r_1a_1)(1 - \frac{1}{R_0})}{\gamma_1 + b_1 - r_1a_1}, R_1 \rightarrow \frac{K_1\gamma_1(1 - \frac{1}{R_0})}{\gamma_1 + b_1 - r_1a_1}$ , so  $P_2$  is global attractive (Fig. 26.2).

To show the global stability of the system around  $P_2$  we choose  $\alpha_1 = 0.13$ ;  $\beta_{11} = 0.2$ ;  $\beta_{12} = 0.1$ ;  $\gamma_1 = 0.2$ ;  $d_1 = 0.12$ ;  $a = 0.45$ ;  $a_4 = 100$ ;  $a_5 = 50$ ;  $\eta = 30$ ;  $\delta_1 = 0.005$ ; So when  $\frac{d_2}{kaK_1} \geq 1, R_0 = 363.3 > 1$ ,  $P_2$  is globally asymptotically stable.

(4) To show the locally asymptotically of the equilibrium  $P_3$ , we use the method of the third approximation. From Theorem 1 and  $\frac{d_2}{kaK_1} < 1$ , easy to know  $N_1 \rightarrow N_{1E}, N_2 \rightarrow N_{2E}$  as  $t \rightarrow \infty$ , also  $N_1(0) \neq 0, N_2(0) \neq 0$ , so progressive autonomous system of (26.2):

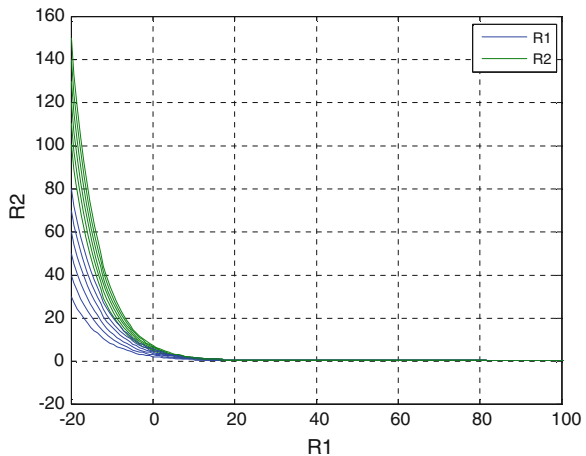
**Fig. 26.2** The global stability of the system around  $P_2$



$$\begin{cases} \dot{I}_1 = (\beta_{11}I_1 + \beta_{12}I_2)(N_{1E} - I_1 - R_1) - \gamma_1 I_1 - \left[ d_1 + (1 - a_1) \frac{r_1 N_{1E}}{K_1} \right] I_1 - a N_{2E} I_1 \\ \dot{R}_1 = \gamma_1 I_1 - \left[ d_1 + (1 - a_1) \frac{r_1 N_{1E}}{K_1} \right] R_1 - a N_{2E} R_1 \\ \dot{I}_2 = (\beta_{21}I_1 + \beta_{22}I_2)(N_{2E} - I_2 - R_2) - d_2 I_2 - \gamma_2 I_2 - n N_{2E} I_2 \\ \dot{R}_2 = \gamma_2 I_2 - d_2 R_2 - \delta_2 R_2 - n N_{1E} R_2 \end{cases}$$

To show the global stability of the system around  $P_3$  we choose  $r_1 = 0.15$ ;  $K_1 = 1000$ ;  $a_1 = 0.13$ ;  $\beta_{11} = 0.2$ ;  $\beta_{12} = 0.1$ ;  $\gamma_{11} = 0.2$ ;  $d_2 = 0.8$ ;  $a_4 = 100$ ;  $a_5 = 50$ ;  $\eta = 30$ ;  $\delta_1 = 0.005$ ;  $k = 0.8$ ;  $a = 0.4$ ; As show in Fig. 26.3, we claim that for  $\frac{d_2}{kaK_1} < 1, R_1 = 0.862 < 1, R_2 = 0.3013 < 1$  the system approaches  $P_3$ . From corollary 1, we can obtain that the bounded solution of (26.2) will approach to global asymptotic stability of the (26.3)'s equilibrium. And

**Fig. 26.3** The global stability of the system around  $P_3$



$I_1 \rightarrow 0, R_1 \rightarrow 0, I_2 \rightarrow 0, R_2 \rightarrow 0$  as  $R_1 < 1, t \rightarrow \infty$ , so  $P_3$  is locally asymptotically, therefore  $P_3$  is globally asymptotically stable. by the same token can prove (5).

**Acknowledgments** This paper is sponsored by the National Natural Science Foundation of China (NSFC, Grant U1204703, 61142010), the Key Scientific and Technological Prject of Henan Prince (112400450505, 122102310004, 122400450544), the Fundamental Research Funds for the Central Univerties (HUST:2012QN087, 2012QN088), the Innovation Scientists and Technicians Troop Construction Projects of Zhengzhou City (10LJRC190), the Scientific and Technological Project of Zheng zhou City (121PRKXF658-4).

## References

1. Yoichi Enatsu YN, Yoshiaki M (2012) Global stability of SIRS epidemic models with a class of nonlinear incidence rates and distributed delays. *Acta Mathematica Scientia* 32B(3):851–865
2. Wang JL, Yasuhiro T (2012) A multi-group sveir epidemic model with distributed delay and vaccination. *J Math Biol* 5(3):1–18
3. Gang Y, Wang LH (2009) A class of the combination of SEIR and SEIS epidemic model with constant recruitment. *J Hennan Polytech Univ (Natural Science)* 28(2):256–259
4. Xiao JY, Wu CF, Weng PX (2012) Traveling waves for a sirs model with nonlocal diffusion. *Int J Biomath* 5(5):1–36
5. Wang LD (2005) Analysis on a SIRS epidemic model with a nonlinear incidence rate. *J North China Inst Technol* 26(1):1–5 (in chinese)
6. Wang J, Wang H (2002) SIRS epidemiological models with population dynamics and nonlinear incidence. *J Ianzhou Univ (Natural Sciences)* 38(1):22–29
7. Xiao YN, Chen LS (2003) Stabilizing effect of cannibalism on a structured competitive system. *Chin J Math Phys* 22A(2):210–216 (in chinese)
8. Herbert WH, Wang WD, Han LT (2004) A predator–prey model with infected prey. *Theor Popul Biol* 66:259–268
9. Cao Y, Jing YW, Yuan F (2012) Analysis of a SIRS model with nonlinear incidence rate on complex networks. *J Northeast Univ (Natural Science)* 33(1):1720 (inchinese)
10. Mukherjee D (2010) Hopf bifurcation in an eco-epidemic model. *Appl Math Comput* 217(5):2118–2124



# Chapter 27

## An Entity Answer Ranking Method Based on MLNs

Fangqiong Chen, Zhengtao Yu, Jianyi Guo, Tao Shen  
and Yantuan Xian

**Abstract** For the characteristics of factoid and list answers in domain Q&A system, we built a ranking model combined with multiple features of domain entity answers based on MLNs. This method uses predicate formulas to describe the relevant features of the questions–candidate answers and the answers–knowledge base, merging these features into Markov Logic Network, and then adopting discriminant training learning algorithm to learn the weights of feature parameters, which can give different weights according to the relevance of different features, finally it use MC-SAT algorithm reasoning to get the relevance of questions and answers, realize the answer ranking. Experiments show that the proposed method can greatly improve the answer precision and recall rates compared with other methods.

**Keywords** Question-answering system · Markov logic network · Learning to rank · Multiple features

### 27.1 Introduction

At present, there are a lot of researches in answer ranking, in TREC2002 QA tasks, Xu used empirical formula to integrate the factors involved in the process of answering the question, which takes the verb, the matching of context and questions as well as the types of questions into consideration, and gives the same

---

F. Chen · Z. Yu (✉) · J. Guo · Y. Xian  
The School of Information Engineering and Automation, Kunming University of Science and Technology, 650051 Kunming, China  
e-mail: ztyu@hotmail.com

Z. Yu · J. Guo · T. Shen · Y. Xian  
The Intelligent Information Processing Key Laboratory, Kunming University of Science and Technology, 650051 Kunming, China

weight for these three factors, thereby resulted in confidence scores [1]. You proposed a model based on maximum entropy algorithm for computing answer confidence with a number of elements taken into account to improve the accuracy of answers. First, extract the following four factors from the training corpus: the answer score, number of steps, the best number of votes and the best vote to train the maximum entropy model; after that apply the trained model on the test set to calculate confidence level; Finally make the correct answer stand in the front as much as possible, in TREC 2002, QA system uses the algorithm to compute the confidence for each answer, and ranks them, which obtains very good results [2]. Sun took the problem of extracting the answer as a classification problem, considering the candidate answer as correct or wrong. They extract the features of questions and candidate answers to train the maximum entropy model, and then use the trained model to extract the answer [3].

The above answer ranking methods do not consider the relationship between data to the answer ranking model, however, there are a variety of relationships between these data in reality, such as similarity, distance, link, and the reference relationships. According to these “relationship”, Markov logic network combines the Markov network with first-order logic, which introduces domain knowledge through the first-order logic into the Markov network, and provides the ability to deal with uncertainty for first-order logic. Markov logic network was first proposed by Richardson [4] in 2004, followed by the study of Domingos, Kok, Singla, et al. The discriminative training for weight learning method and the structure learning method for the optimization of type of likelihood methods were proposed, but it is still a new study in Artificial Intelligence, for learning and inference algorithm is very limited. In 2008, Wu published an article on the WWW, he extracted information by Markov logic network algorithm, and proposed a refining the Wiki knowledge automatically method, which effectively simplifies the manual workload [5]. In 2009, Poon took Markov logic network in unsupervised semantic analysis, and achieved fairly good results [6]. Using Markov logic network can take advantage of “a variety of relationships” in different areas to effectively integrate correlation algorithms to obtain better results. Based on all kinds of “relationship” features in the question answering system, this paper proposes a new research approach that Markov logic network is applied to combine a variety of relationships to solve the ranking problem in the question answering system, and builds the answer ranking model.

## **27.2 The Feature Extraction of Questions and Answers**

### ***27.2.1 Context Features***

Context has different meanings in different fields. In the field of information, context is defined for the collection of attributes describing the semantic similarity between objects [7]. The paper extracted from the context information between the

questions and answers, the definition of context features between questions and answers is as follows:

### 27.2.1.1 Lexical Features

Feature words vector represents the relative order of the feature words set and feature words. The two sentences contain the same words, there is some correlation between these two statements. The TF-IDF based on vector space is one of the common methods [8]. In addition, containing the same words in the consistency of the order is also reflects the similar degree of two sentences.

Defined questions  $S_q = (A, B, C)$ , the candidate answers  $S_a = (D, C \text{ and } A \text{ and } C)$ , the letter in parentheses represents the feature words of the statement.

The definition of feature words vector similarity is as follows:

$$S_w = \frac{2 * \sum_{W_i \in S_q} \lambda_w}{|S_q|(|S_q| + |S_a|)} |Length(S_a, S_q)|$$

where  $W_i$  represents the feature words in the candidate answers,  $\lambda_w$  indicates the weight of the same words, the same words only counts once here.  $|S_q|$  and  $|S_a|$  respectively express the dimension of the feature word vectors of the questions and candidate answers.  $|Length(S_a, S_q)|$  signifies the number of elements of the longest fragment which candidate answers match the feature words set of question.

### 27.2.1.2 Semantic Similarity Features

Except the statistical characteristics of the term in the context, the semantic information of the term itself is also an important feature. With the help of HowNet, we compute the semantic distance between questions and answers as the semantic similarity integrated into the ranking model [9].

## 27.2.2 Density and Frequency Features

This evaluation indicator measures the density and frequency of question terms in the answer text, mainly includes: numbers, distances, word sequences, etc. Variants of these features were used previously for answer extracting in factoid QA [10]. We will discuss briefly in the following:

Quantity features: The matching nouns (verb, numeral or quantifier) of candidate answers account for the proposition of nouns (verb, numeral or quantifier) in query words. The features reflect the similarity of designated part of speech and questions in word matching level.

Distance Features: The distance of candidate's answers from the words or phrase of questions. Compute the distance for each sentence contained the candidate answers and yield a weighted mean value.

$$DS = avg \sum_i \frac{1}{abs(pos(keyword_i) - pos(answer))}$$

Sequence Features: This feature investigate whether the matching words of sentence, answer sequence and questions sequence are in the same order, measured by the words in same sequence percent of the question words number.

### 27.2.3 External Knowledge Features

We get the Boolean features values of relevant answers by validating the online knowledge base, if the answers appear in Wikipedia or Baidu encyclopedia, the corresponding feature value is 1, otherwise, the value is 0.

## 27.3 Construct the Entity Answer Ranking Model Based on MLNs

Implementing the model is to introduce the features relationship into Markov logic network, learning these features to get the relevance of questions and answers, specifically through the following three steps to achieve: predicate and formula definition, weight learning, and probabilistic reasoning algorithm.

### 27.3.1 Predicate and Formula Definition

Due to the ranking task is essentially to rank the candidate answers, usually through the relevant features of the questions and answers to get the relevance. In Markov logic network the features and relativities between the questions and answers are expressed by the predicate, give an example to illustrate:

- (1) SimilarityWord (question, answer) denotes: feature word vector similarity of question and answer;
- (2) Distance (question, answer) denotes: distance of question and answer.

According to these predicates, we can get the formulas:

- (1) SimilarityWord (question, answer) => related (question, answer), indicates that if feature word vector similarity is more than 0.1, then the question and answer is related;

- (2) Distance (question, answer) => related (question, answer), indicates that if distance is less than a certain value, we set the value as 20, then the question and answer is related.

For the different features make different contribution to the questions and answers, the formula of Markov logic network is to add “+” sign, these formulas can be respectively learned different weights based on variables represented by different individuals. In mln file a formula is written as + Similarity (question, answer), that for a pair of question and answer, to learn the weights between the different features and relevance of answers.

### 27.3.2 Probabilistic Reasoning Algorithm

The basic task of the reasoning is that given the evidence of text set to find a most possible world  $y$ . That is to solve the probability of established formula  $F_1$  under the conditions of given formula  $F_2$  Markov logic network is able to answer this question. The task of reasoning in this paper is that given relevant features between questions and answers, in order to determine the relevance of the questions and answers.

In order to achieve better results, we first construct the minimum closed network of query predicate. The construction process of this network is as follows:

- (1) View if the evidences contain the predicate of the current query predicate depending on. If so, skip to (2), did not skip to (3);
- (2) Detect whether the query predicate set has been checked up. If so, then the algorithm terminates; if not, perform a query predicate, and skip to (1);
- (3) Join predicate  $C$  to the network;
- (4) Check the predicate that is dependent on predicate in (3) whether exists in the evidence set. If so, skip to (2), did not skip to (3).

After constructing a query predicate minimum closed network, we have to sample for all nodes in the network through the MCMC reasoning algorithm, this paper use the MC-SAT algorithm. The algorithm is to construct reversible Markov transfer nuclear through a pre-defined invariant distribution, and in essence, it is an auxiliary variable method. Compared with the standard MCMC approach, the rate of convergence of this algorithm is much faster.

Algorithm steps are as follows:

- (1) Initialize the set of states that meets the strict clause;
- (2) Sum all closed clauses that meet states in step (1) to the clause set  $M$  by the probability of  $1 - e^{-w}$ ;
- (3) Extract a sample evenly from all states that meet  $M$  and then turn to (2) until the number of all samples has been finished.

After extracting  $n$  samples, the expectation is  $E_{w,y}[n_i(x, y_w^*)] \approx \sum_{l=1}^n \frac{n_i(x^{(l)})}{n}$ , that is, the expectation of the function is approximated as the mean of the sample.

## 27.4 Experiment and Comparative Analysis

The evaluation methods of the factoid question and the list question are different, so we took separate experiments and comparative analysis. For the factoid question, two evaluation criteria were used: the average accuracy rate of MRR (Mean Reciprocal Rank) and Recall (Recall). For the list question, in addition to the above two evaluation, F value (F-measure) is also the evaluation criteria.

### 27.4.1 Dataset

At present, there is no very authoritative corpus in Chinese Q&A system. For the field of tourism in Yunnan, we have separately collected 150 factoid questions and list questions for attractions name, place name, hotel name, traditional festival, local snack and other entities. We use entity recognition model to make entity recognition for answer candidate document. From the 300 entities answer of artificial screening, on the basis of the relevance of the questions and answers, the relevant was labeled as (2), the part relevant was labeled as (1), the irrelevant was labeled as (0). Thereby the available corpus resources were obtained. Table 27.1 shows the corpus.

The experimental dataset is divided into training set, validation set and test set. The training set is used to train ranking model, the validation set is used to adjust the parameters of the ranking model, and the trained ranking model will be tested for the performance of answer ranking on the test set.

### 27.4.2 Experiments and Results Analysis

First, analyze the questions, then, research and select features of questions and candidate answers, and the candidate answers and online knowledge base.

**Table 27.1** Answer ranking corpus

	Total	Attractions	Place	Hotel	People and culture	Other
Factoid question	150	35	35	30	30	20
List question	150	35	35	30	30	20

Secondly, integrate the features relationship into Markov logic network to construct ranking model. Thirdly, figure out predicate weights by structure learning, parameter learning and reasoning algorithm. Finally, the trained ranking model will be tested on test set.

### 27.4.2.1 Comparison with Other Ranking Methods

In order to evaluate the performance of answer ranking model based on MLNs, we take respectively answer linear score and BM25 as benchmark to do experiment comparison. In the experiment, we use NDCG@n to measure ranking results, the comparison results are shown in Table 27.2.

By the above experiment can be seen, answer extracting method based on MLNs is obviously better than linear score and BM25. Because of the dispersion of the list question answers distribution, its top accuracy rate is low compared to factoid question. When N increased to 3, the NDCG @ N value of ranking model based on MLNs did not change, the other two algorithms had declined, which indicated that ranking model based on MLNs can ensure accuracy rate for list questions.

### 27.4.2.2 Comparison of Answer Extraction Efficiency

In order to measure the contribution of the answer ranking model for answer extraction, we will compare the answer ranking model based on MLNs with logistic regression and maximum entropy model. We are more concerned about the accuracy of correct answers ranked top-1, so we adopt the mean reciprocal rank (MRR) and the recall rate (R) and F value (F-measure) to evaluate. The results are shown in Table 27.3.

Experimental result shows that the proposed method improved the accuracy and recall rate greatly compared to traditional ranking model. The ranking model based on MLNs can effectively integrate the various features relationship between the questions and answers, and answers and knowledge base, and get different weights of various features on the contribution for the correct answer via parameter learning and reasoning. It effectively improved the performance of answer ranking to some extent, compared to the maximum entropy model and logic regression model.

**Table 27.2** The comparison of different answer ranking methods

Ranking method	NDCG@1		NDCG@2		NDCG@3	
	Factoid	List	Factoid	List	Factoid	List
Linear score	0.73	0.72	0.71	0.69	0.68	0.67
BM25	0.75	0.73	0.73	0.71	0.69	0.70
MLNs	0.78	0.77	0.76	0.75	0.76	0.75

**Table 27.3** The comparison of different answer extracting methods

Answer extracting method	Question type	Evaluation index		
		MRR	Recall	F-measure
Maximum Entropy	Factoid	0.48	77.5 %	76.3 %
	List	0.36	70.4 %	72.4 %
Logistic regression	Factoid	0.58	82.8 %	86.5 %
	List	0.40	78.6 %	78.3 %
MLNs	Factoid	0.60	89.8 %	89.9 %
	List	0.44	89.4 %	87.4 %

## 27.5 Conclusions

For the characteristics of factoid and list answers in domain Q&A system, A ranking model was built which combined with multiple features of domain entity answers based on MLNs. The construction method of entity answer ranking model based on MLNs will take feature relationship of questions and candidate answers, the candidate answer and Knowledge base into the ranking model according to the different weights to improve the answer precision. On this basis, doing research on answer ranking for the factoid and list questions from the recall rates and precision to improve answer ranking results, and the experiments show that the proposed method can effectively improve the answer precision and recall rates.

**Acknowledgments** This paper is supported by National Nature Science Foundation (No. 61175068), and the Open Fund of Software Engineering Key Laboratory of Yunnan Province (No. 2011SE14), and the Ministry of Education of Returned Overseas Students to Start Research and Fund Projects.

## References

1. Xu J, Licuanan A, May J, et al (2002) TREC 2002 QA at BBN: answer selection and confidence estimation. In: Proceedings of the 11th text retrieval conference (TREC-11). NIST Special Publication, Gaithersburg, pp 96–101
2. You L, Zhou Y-Q, Huang X-J et al (2005) QA system based on maximum entropy model confidence score algorithm. *J Softw* 16(8):1407–1414
3. Sun A, Jiang M-H, Ma Y-J (2006) An instance-based approach for pinpointing answers in chinese question answering. In: Proceedings of 2006 international conference on signal processing (ICSP 2006)
4. Richardson M, Domingos P (2006) Markov logic networks. *Mach Learn* 62:107–136
5. Wu F, Weld D (2008) Automatically refining the wikipedia infobox ontology. In: Proceeding of the 17th international conference on world wide web. ACM, New York, pp 635–644
6. Poon H, Domingos P (2009) Unsupervised semantic parsing. In: Proceedings of the 2009 conference on empirical methods in natural language processing. ACL, Singapore
7. Davis J, Domingos P (2009) Deep transfer via second-order Markov logic. In: Proceedings of the 26th international conference on machine learning. ACM Press, Montreal



8. Tian Y-H (2011) Research on context-based statistical relational learning. Dissertation, Institute of Computing Technology, The Chinese Academy of Sciences, Beijing
9. Hermjakob U (2002) A question answering system based on vector similarity. In: Proceedings of the 26th annual international ACM SIGIR conference, USA, pp 126–132
10. Liu Q, Li S-J (2002) Word similarity computing based on How-net. In: Proceedings of the 3rd Chinese lexical semantics, 2002–2005

# Chapter 28

## A Chinese Expert Name Disambiguation Approach Based on Spectral Clustering with the Expert Page-Associated Relationships

Wei Tian, Tao Shen, Zhengtao Yu, Jianyi Guo and Yantuan Xian

**Abstract** Aimed at the problems of Chinese experts' name repetition and representation diversity, a Chinese expert name disambiguation approach based on spectral clustering with the expert page-associated relationships is proposed. Firstly, the TF-IDF algorithm is used to calculate the word-based feature weights, and then the cosine similarity algorithm is employed to compute the similarity between the evidence-pages to obtain the initial similarity matrix of expert evidence-pages. Secondly, the expert page-associated relationship features are taken as the semi-supervised constraint information to correct the initial similarity matrix, and next the spectral clustering-based method is used to build expert disambiguation model. Finally, taking the contrast experiments on Chinese expert evidence-page corpus of manually labeled, the result shows that the semi-supervised spectral clustering on Chinese experts' name disambiguation method with the expert page-associated relationships than that without the associated constraint information, the F-value has an average increase of 9.02 %.

**Keywords** Page-associated relationships · Chinese expert name disambiguation · The spectral clustering · Semi-supervised constraint information

---

W. Tian · T. Shen (✉) · Z. Yu · J. Guo · Y. Xian

The School of Information Engineering and Automation, Kunming University of Science and Technology, NO. 727 South Jingming Road, Chenggong District, Kunming 650500, Yunnan, People's Republic of China  
e-mail: shentao1@gmail.com

T. Shen

The School of Material Science and Engineering, Kunming University of Science and Technology, NO. 727 South Jingming Road, Chenggong District, Kunming 650500, Yunnan, People's Republic of China

## 28.1 Introduction

With the rapid development of the Internet, the name repetition on the Internet has become one of the most common phenomena, which brought great difficulties to the Expert search and experts' resource utilization, therefore, expert name disambiguation is imperative. The expert disambiguation methods existing are as follows: The first kind is a similarity calculation-based clustering disambiguation method, such as Wang proposed to use the vector space model of web content to do expert evidence-pages clustering disambiguation to solve the multi-document coreference resolution problem [1]. Bollegala put forward the experts clustering disambiguation solution on key phrases extraction automatically in the context and computing similarity [2]. The second is a hierarchical-based clustering disambiguation approach, such as Zhang used hierarchical clustering algorithm to solve the multi-document ambiguity issue of Chinese names [3]. The third is a clustering disambiguation method based on the specific relationships. For example, Lang presented a name disambiguation approach based on social networks, exploited the evidence-page titles and name co-occurrence relationships in the context fragment to build social networks and used clustering to realize disambiguation [4]. Tang in Tsinghua University proposed the clustering disambiguation combined with the attributes of expert papers and paper-cooperative relations, selected the article title, abstract, author, etc. as features with published paper-cooperative relations, through HMRFs [5] (Hidden Markov Random Fields) clustering method to do expert clustering disambiguation [6]. HMRFs itself can't avoid order-dependence issue. Given the current cluster centroids, the order in which points are greedily reassigned to clusters determines the new clusters.

Some above also used many different relationships to realize the expert name disambiguation, but did not consider fully the relationships between the pairwise expert evidence-pages, while many associated relationships existing in the content of a large number of expert evidence-pages have a significant supporting role on expert identity determination. According to the attribute characteristics of Chinese experts, proposes a Chinese expert name disambiguation approach based on spectral clustering with the expert page-associated relationships.

## 28.2 Building the Model of Chinese Expert Disambiguation with the Page-Associated Relationships

Shi and Malik proposed a graph partitioning criterion—the normalized cut [7], and made which optimization in order that the nodes have high similarities within the cluster, and low similarities between the clusters.

Primarily, use the cosine similarity to calculate the similarity between the expert evidence-pages, and construct the initial similarity matrix, and then adopt the expert page-associated relationships as semi-supervised constraint information to adjust

the similarities of the expert evidence-pages, and then correct the initial similarity matrix. Next, build the corresponding Laplace matrix and then make the Laplace matrix normalized. Finally do spectral feature decomposition on the obtained normalized matrix, and then cluster based on the spectral methods. The following details the building process of the spectral clustering model of Chinese expert disambiguation with the page-associated relationships.

### 28.2.1 Constructing the Initial Similarity Matrix of the Expert Evidence-Pages

Expert name disambiguation is to make the evidence-pages of different experts in the same name classified. Further more, the disambiguation process can be regarded as the clustering procedure of the expert evidence-pages. First of all, taking the expert evidence-page collection that needs disambiguation as the graph  $G = (V, E, A)$ , where  $V$  is the node set of the expert evidence-pages,  $A$  indicates the similarity matrix between the evidence-page nodes,  $E$  is the edge set. Assign a weight to the edge between two nodes according to the similarity between them. After some pretreatment for the retrieved expert evidence-pages, such as stripping tags, Chinese word segmentation and the stop word removal, we entirely consider the occurrence frequency of different words in all the texts and the resolution capabilities of the word to the different texts, and then use the TF-IDF based on word frequency statistics to calculate the word feature weights of the two evidence-page nodes. Take the form of vectors to express a document, and each vector is represented by the feature item and its corresponding weight, thus a document vector space is constituted. The cosine of the vector space angle of the two documents is employed for defining the similarity between the two expert evidence-page nodes, based on which we can get the initial similarity matrix  $A$ .

Suppose there are two arbitrary evidence-page nodes  $x_i, x_j \in V$ , TF-IDF formula is:

$$W_{t,x} = TF_{t,x} \times IDF_{t,x} \quad (28.1)$$

$$TF_{t,x} = \frac{N}{M} \quad (28.2)$$

$$IDF_{t,x} = \log(X/X_N) \quad (28.3)$$

where  $W_{t,x}$  is the weight of the feature item  $t$  in the document  $x$ ;  $TF_{t,x}$  represents the occurrence frequency of  $t$  in the document  $x$ . Formula (28.2) shows that there are  $N$  same feature items in the document which contains  $M$  words, and this formula can mirror the distribution of the feature item inside the document.  $IDF_{t,x}$  is known as the document frequency of features to reflect the distribution of feature item  $t$  in the whole document set and the distinction ability of this feature item to

a certain extent, and where  $X$  devotes the number of all the documents in the document set,  $X_t$  represents the occurrence frequency of  $t$  in the document set.

The following we use the angle cosine of the two document vectors to define the initial similarity  $A_{ij}$  between two evidence-page nodes:

$$A_{ij} = Sim(x_i, x_j) = \cos \theta = \frac{\sum_{t=1}^n W_{t,x_1} \times W_{t,x_2}}{\sqrt{\left(\sum_{t=1}^n W_{t,x_1}^2\right) \left(\sum_{t=1}^n W_{t,x_2}^2\right)}} \quad (28.4)$$

where  $W_{t,x_1}, W_{t,x_2}$  is respectively devotes the weight of the feature item  $t$  in the evidence-page document  $x_1, x_2$ . And further obtain the initial similarity matrix  $A$ .

### 28.2.2 Correcting the Initial Similarity Matrix Based on the Page-Associated Relationships

#### 28.2.2.1 Selecting the Page-Associated Features

On the basis of the expert evidence-page content, define the expert evidence-page-associated features firstly. The features are defined in Table 28.1.

#### 28.2.2.2 Page-Associated Feature Constraints and Similarity Correction

Assume two arbitrary evidence-page nodes  $x_i, x_j \in V$ , define a adjustment matrix  $A'_{ij}(i, j = 1, 2, \dots, n)$  as follows:

$$A'_{ij} = \sum_{m=1}^7 \alpha_m f_m + \alpha_8 f_8 \quad (28.5)$$

**Table 28.1** The expert page-associated features

Serial number ( $m$ )	Feature name	Feature type	Eigenvalue ( $f_m$ )	Feature weight ( $\alpha_m$ )
1	Page title association	Boolean	0,1	$\alpha_1$
2	Links pointed to each other	Boolean	0,1	$\alpha_2$
3	Whether the same education	Boolean	0,1	$\alpha_3$
4	Organization co-occurrence	Boolean	0,1	$\alpha_4$
5	Research association	Boolean	0,1	$\alpha_5$
6	Whether the same gender	Boolean	0,1	$\alpha_6$
7	Whether the same place of birth	Boolean	0,1	$\alpha_7$
8	Whether the same date of birth	Boolean	0,1	$\alpha_8$

where  $\alpha_m$  is the page-associated feature weight obtained by training. Therefore, set  $\alpha_8 = 0.5$ ,  $\sum_{m=1}^7 \alpha_m = 0.5$ , and  $f_m$  devotes the page-associated eigenvalues mentioned in Table 28.1.

Wagstaff first introduced two types of pairwise constraints in literature [8], that is, use the “must-link” and “cannot-link” to assist the cluster search. Define the page-associated feature constraints (including the “must-link” and “cannot-link”) shown in formula (28.6) as the supervision constraint information, and then do the correction of the initial similarity matrix  $A$  based on the “must-link” and “cannot-link” constraints to obtain the final similarity matrix  $A''$ .

$$\begin{cases} A''_{ij} = A''_{ji} = 1, (A'_{ij} \geq 0.5) \Leftrightarrow (x_i, x_j) \in \text{must-link} \\ A''_{ij} = A''_{ji} = 0, (A'_{ij} = 0) \Leftrightarrow (x_i, x_j) \in \text{cannot-link} \\ A''_{ij} = A_{ij}, (0 < A'_{ij} < 0.5) \end{cases} \quad (28.6)$$

For formula (28.6), it is represented that it has the same date of birth at least or that the previous seven eigenvalues are all 1 in the two evidence-pages if  $A'_{ij} \geq 0.5$ , so that the two evidence-pages together are a “must-link” constraint. The same reason is that all page-associated eigenvalues are all 0 when  $A'_{ij} = 0$ .

### 28.2.3 Constructing the Model of Chinese Expert Name Disambiguation

Here are the steps of the spectral clustering method with the page-associated feature constraints:

1. Initialization firstly. For  $\forall x_i, x_j \in V$ , compute the initial similarity between the evidence-page nodes to get the initial matrix  $A$  as shown in formula (28.4).
2. Define the page-associated features, calculate the eigenvalues based on word matching, then train to obtain the corresponding feature weights to get the adjustment matrix  $A'_{ij} = \sum_{m=1}^7 \alpha_m f_m + \alpha_8 f_8$ , where  $\alpha_8 = 0.5$ ,  $\sum_{m=1}^7 \alpha_m = 0.5$ .
3. Define the page-associated feature constraints according to the adjustment matrix  $A'$  as shown in formula (28.6), obtain the corrected similarity matrix  $A''$  after judgment and adjustment.
4. Define the diagonal matrix  $D_{ii}$  of  $A''$ :  $D_{ii} = \text{diag}(d_1, d_2, \dots, d_n) = \sum_{j=1}^n A''_{ij}$ .
5. Construct the Laplace matrix  $L = D - A''$ .
6. Normalize:  $N = (d_{\max} I - L) / d_{\max}$ , where  $d_{\max}$  is the maximum row sum of  $A''$ .
7. Find  $y_1, y_2, \dots, y_k$ , the  $k$  largest eigenvectors of  $N$  and form the matrix  $Y = [y_1, y_2, \dots, y_k] \in \mathbb{R}^{n \times k}$ , where  $y_i (i = 1, 2, \dots, k)$  is the column vector.
8. Normalized the rows of  $Y$  to get the matrix  $Z$ , where  $Z_{ij} = Y_{ij} / \left( \sum_j Y_{ij}^2 \right)^{\frac{1}{2}}$ .

9. Treating each row of  $Z$  as a point in  $R^k$ , cluster into  $k$  clusters using K-means.
10. Assign the original point  $x_i$  to cluster  $j$  if and only if row  $i$  of  $Z$  was assigned to cluster  $j$ .

## 28.3 Experiments and Analysis

### 28.3.1 Experiment Data Preparation

Do the spider based on CNKI resources for 2000 entity names of Chinese expert engaged in the field of computer information processing, and there are 2000 expert evidence-pages in total. The expert evidence-pages include the Expert Homepage, Baidu baike, Soso baike, HDWiki, Wikipedia, MicroBlog, Boke and so on. There are four selections of test sets in the experiments. Randomly chose 80 expert names, that is, a total of 800 evidence-pages from the original data set are considered as a test data set each time. The number of page-associated feature constraints (including the “must-link” and “cannot-link”) is between 0 and 300 derived from the experiments. The page-associated feature constraints are randomly generated from the training data set.

### 28.3.2 Experimental Evaluation

Here the F-value will be employed. The F-value is defined as below:

$$Pre = \frac{T_p}{T_p + F_p}, \quad Rec = \frac{T_p}{T_p + F_n}, \quad F = \frac{2Pre \times Rec}{Pre + Rec} \quad (28.7)$$

where  $T_p$  is the number of evidence-page documents that the two documents together in one cluster are classified correctly,  $T_n$  refers to the number of evidence-page documents that the two should not be gathered in one cluster are properly separated,  $F_p$  represents the number of evidence-page documents that the two should not be placed in one cluster are divided into one falsely,  $F_n$  is the number of evidence-page documents that the two should not be separated are parted wrongly.

### 28.3.3 Experiments and Results Analysis

Experiment 1 is to test the impact on the clustering of inputting different numbers of associated feature constraints. As the final clustering performance result with page-associated constraints, it would be averaged over the clustering results of

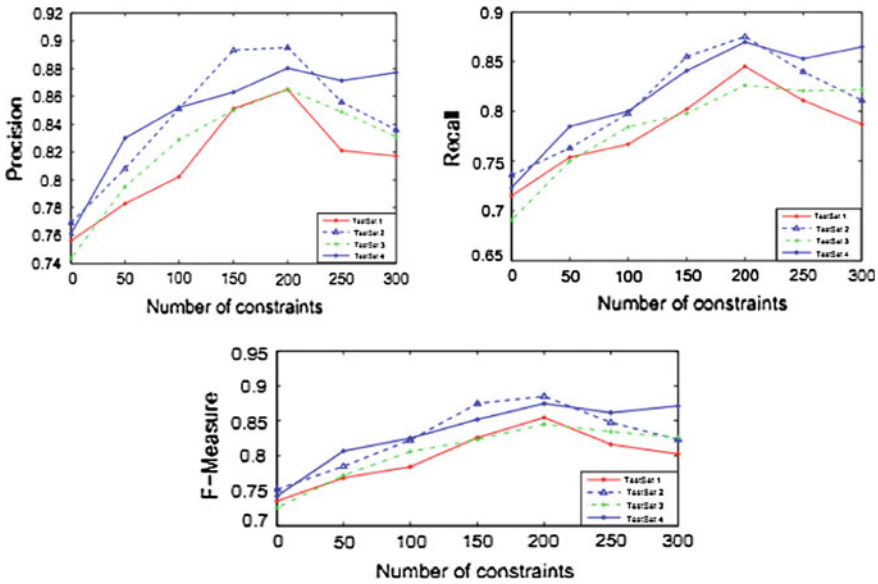


Fig. 28.1 The impact of different numbers of associated constraints on the P, R and F-value indexes of the four test sets

each test set. As the number of constraints is continuous, it is not convenient to use tabular data to show out the results. So Fig. 28.1 reveals the changes of the P, R and F-value indexes on the four test sets with the different numbers of associated constraints.

It can be seen from Fig. 28.1 that the P, R and F-value indexes obtained by our approach show upward trend as a whole with the increase in the number of constraints. When the number of constraints is between 150 and 200, the corresponding P, R and F-values reached the maximum and the number of clusters obtained at this time is optimal. The clustering performance appears downward trend when the number of constraints is greater than 200.

Experiment 2 is the clustering performance comparison of the spectral clustering method with the expert page-associated constraints and without constraints. The P, R and F-values, which are acquired by averaging over the data in Fig. 28.1, represent the final results of the spectral clustering approach with the page-associated constraints. There are the results of Experiment 2 given in Table 28.2.

Seen from Table 28.2, the P, R and F-values, obtained by using the spectral clustering method with the expert page-associated feature constraint, are significantly improved than that without the constraint information. Therefore, it can be known that adding constraint information can largely enhanced the clustering performance of the entire data set.



**Table 28.2** Indicator changes before and after adding the page-associated constraints

Sets Methods	The spectral clustering method without the page-associated constraints			The spectral clustering method with the page-associated constraints (average value)		
	P (%)	R (%)	F (%)	P (%)	R (%)	F (%)
TestSet 1	75.61	71.50	73.50	82.32	79.43	80.92
TestSet 2	76.91	73.57	75.21	85.64	82.35	83.97
TestSet 3	74.38	69.07	72.55	83.65	80.01	81.78
TestSet 4	76.13	72.35	74.19	86.17	83.57	84.86

## 28.4 Conclusions

The expert page-associated features, considering as supervision constraints, are introduced into the spectral clustering algorithm. We propose a Chinese expert name disambiguation approach based on spectral clustering with the expert page-associated relationships. The comparative experiment has fully demonstrated the validity of expert page-associated features viewed as priori constraints and their positive role in guiding the spectral clustering algorithm. Our further work is how to actively provide the constraint information with rich content for the spectral clustering in order to utilize these limited constraints to significantly improve the performance of the clustering disambiguation.

**Acknowledgments** This paper is supported by National Nature Science Foundation (No. 61175068), and the Open Fund of Software Engineering Key Laboratory of Yunnan Province (No. 2011SE14), and the Ministry of Education of Returned Overseas Students to Start Research and Fund Projects.

## References

1. Wang H, Mei Z (2005) Chinese multi-document person name disambiguation. *High Technol Lett* 11(3):280–283
2. Danushka B, Yutaka M, Mitsuru I (2006) Disambiguation person names on the web using automatically extracted key phrases. In: Brewka G, Coradeschi S, Perini A, Traverso P (eds) *Proceedings of the 17th European conference on artificial intelligence*. IOS Press, Riva del Garda, Italy, pp 553–557
3. Zhang S, You L (2010) Chinese people name disambiguation by hierarchical clustering. *New Technol Libr Inf Serv* 2010(11):64–68 (in Chinese)
4. Lang J, Qin B (2009) Person name disambiguation of searching results using social network. *Chin J Comput* 32(7):1365–1375
5. Quattoni A, Wang S, Morency LP et al (2007) Hidden conditional random fields. *IEEE Trans Pattern Anal Mach Intell* 29(10):1848–1852
6. Tang J, Yao L, Zhang D et al (2010) A combination approach to web user profiling. *ACM Trans Knowl Discov Data* 5(1):2-2

7. Shi J, Malik J (2000) Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Mach Intell* 22(8):888–905
8. Wagstaff K, Cardie C (2000) Clustering with instance-level constraints. In: Langley P (ed) *Proceedings of the 17th international conference on machine learning*. Morgan Kaufmann Publishers, San Francisco, pp 1103–1110

# Chapter 29

## The Hierarchical Heterogeneous of Parallel Computing Model Based on Method Library

Jibing Duan, Xiaopeng Ji, Jinye Dou and Zhiqiang Wei

**Abstract** This paper puts forward a novel hierarchical heterogeneous of parallel computing model that based on method library. In the original models, although the models provide a model algorithm language for usage, the developers still need to rewrite the methods when calling and the usage is apparently very complex for the developers, demanding the developers' programming skill. In this regard, this paper presents a new hierarchical heterogeneous of parallel computing model which is based on the method library as well as its management system. While developers use the methods, they should only know the parameters of the methods called without having to know the methods of the preparation process, which can help the developers call the method library methods easily to facilitate the preparation of the algorithm program. From experiments results based on the method library, conclusions are drawn. Experiments show that the introduction of the method libraries and its management systems can not only reduce the difficulty of developers, but also reduce compile time and accelerate the operation of the program.

**Keywords** Hierarchical heterogeneous · Layered parallel heterogeneous · Management system · Method library · Model algorithm language · Speed up processing

### 29.1 Introduction

Parallelism has been used widely for many years, mainly in high-performance computing [1], such as petroleum exploration industry. With the development of these different parallel styles, the parallel computing models are established. The

---

J. Duan (✉) · X. Ji · J. Dou · Z. Wei  
Department of Information Science and Engineering, Ocean University of China, Qingdao,  
Shandong, China  
e-mail: yanzhileiliurenzui@gmail.com

existing parallel models are mostly based on the parallel algorithm design and analysis. But most of them have the same defects: poor compatibility, communication overhead, the high coupling degree, the high difficulty on expanding and modifying.

The appearance of the general model of hierarchical heterogeneous can solve the problem conveniently [2]. The general model of hierarchical heterogeneous, divide the parallel tasks into several phases, so that the special personnel can face the different phase to solve the problems. The person, who uses the model to solve the problems, can only divide the problems into several tasks, and then use the new language to code the program, which can be translated to source code. The source code can be built and linked, and the result appears. This process promise the problems solved.

The advantages of the model are: easily used, the low degree of the coupling, the good compatibility, easily expanding. But an obvious disadvantage is that it exists: the high-level language to write code to achieve complex and error-prone. In order to solve the problem, this paper, the idea of a method-based base system to modify the model. According to different tools, according to the method used, or may be used, as well as a commonly used method in industry to achieve deposit method library, so that in use to facilitate the call. This not only reduces the difficulty of program development, and more effective to speed operations.

In the paper, according to the tests of experimental data preprocessing module in the rapid modeling of seismic body, verify the utility of method library in the layered model of heterogeneous parallel.

## 29.2 Related Work

Parallel computing model usually abstracts the basic feature of different parallel computers forming an abstract computational model from the design and analysis of parallel algorithms. In a broader sense, the parallel computing model provide hardware and software interface, and in the interface conventions, the designers of hardware and software in the parallel system can develop support mechanisms for parallelism, thereby improving system properties [3].

The main existing parallel computing models are as follows: PRAM (Parallel Random the Access Machine) model, BSP model (MIMD computing model of distributed memory), LogP model (a multi-processor model features in  $f$  distributed memory and point-to-point communication) BDM (Block Distributed Model), C3 model (the bridge model between the model of shared memory programming and model of distributed memory system based on messaging).

In these universal models, the main factors affecting the efficiency of data access under the environment of shared memory is the lock competition, however, it is the communication cost that affects the efficiency of data access in the multi-machine distributed storage environment. The communication cost is caused not only by the transmission words and transmission time of each word but also by the

cost of starting communication and the communication distance. The more the number of nodes and the longer the communication distance, the greater the system overhead, therefore, while developing the cluster computing technology, we still need to seek ways to enhance the performance of the single node.

Therefore, it is necessary for the emergence of parallel computing model of hierarchical heterogeneous. The parallel computing model of hierarchical heterogeneous consists of three phases: the program model algorithm design phase, the parallel program design phase, and the parallel program execution phase [4, 5].

Figure 29.1 shows the hierarchical model of heterogeneous parallel, in the first phase, program model algorithm on the application layer, reflected on the support layer model algorithm language, the parameter library, after explanation of the interpretation system, the source code appear. And then after the execution of the build system, capable of forming machine language and executes the generated results [6].

In the program model algorithm design phase, the developers are faced with the parameterization of the parallel machine [7]. Developers design the appropriate program model algorithm by analyzing the target. Different from the traditional parallel development, developers need not concern more about the concrete realization of the algorithm at this phase. By interpreting the program model algorithm into different parallel functions based on different hardware and software framework and referring to the software and hardware parameters of current system, overhead function and computing behavior, we can optimize the parallel functions and improve the efficiency of system processing at the same time.

In the parallel program design phase, developers should realize the corresponding parallel program for the various hardware and software interfaces. The parallel part in the program is interpreted as a parallel computing among the compute nodes, a parallel computing of multi-core CPU in the nodes and a parallel computing of multi-core GPU according to the computing scale and problem characteristics.

In the parallel program execution phase, the parallel program is compiled, built, and linked into machine language by the compiler and runs on the corresponding hardware and software framework [8]. For the different computation models and implementation hardware used in the execution phase of parallel program, we

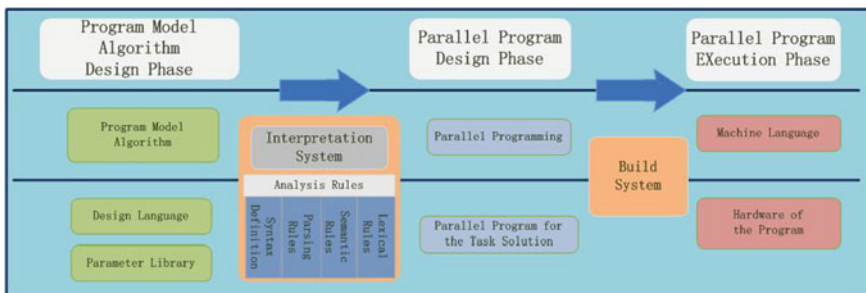


Fig. 29.1 Flowchart of hierarchical heterogeneous model

adopt different compiling systems to compile the program into the machine codes in the parallel program execution level.

At the application level, firstly, application developers should design process model algorithm to describe the target problem; the program model algorithm will finally be interpreted as a serial program and parallel programs based on different hardware framework by the interpretation system; parallel program can be compiled into the corresponding machine code by the different compiling systems and can execute in different hardware frameworks. In the support level, a variety of hardware frameworks (e.g. clusters, multi-core CPU, multi-core GPU) work together to complete the common computing tasks.

## 29.3 Method Library System

### 29.3.1 *Methods Library*

The method used, or may be used, as well as within a particular industry, use the parallel mode, stored in the method base, in order to use conveniently call. Method Library uses the technology of dynamic build [9, 10].

In program model algorithm design phase, the introduction of the method library will be effective enough to reduce the developer's difficult to develop, help reduce compiling time, and accelerate the operation of the program.

The method library is an available ordered collection of a variety of standard procedures, subroutines, and file and its directory information. At the same time, the functions in the method library are standardized. The so-called standardization has triple meanings: the format of the same library of all method programs is uniform; the calling of the functions is the same, the parameter number of each function, the order and the type of each function are in strict rules.

Based on different hardware and software frameworks (for example, MPI, PVM; OpenMP, TBB; CUDA, OpenCL), a parallel task corresponds to a set of parallel functions in method library. The developers match the parallel tasks and the parallel program in method library by calling the command interfaces in method library, and then, the developers can use the instruction to appoint some parallel tasks corresponding to a set of parallel functions in different hardware and software frameworks.

In the current computing model, if we want to introduce a new hardware and software framework, both the original computing model and parallel program should to be modified modify in order to make the new hardware and software framework corporate with the original framework. However, in the common parallel model of the hierarchical heterogeneous, the developers only need to add the parallel program model that corresponds to the hardware and software framework to the method library, and configure the parallel machine parameters to the parameter library, and then we can achieve an extension of the model and does not require any changes to the original parallel modules.

Generally speaking, the introduction of the method library has the following advantages: provides a tool that can calculate, analysis and process; improve the computational efficiency of the process as a whole; achieve code reuse.

### ***29.3.2 Management System of Methods Library***

Sorted by the method function, parallel method library consists of two parts: common library and special library.

Common library provides the interface of parallel functions that used commonly. All these functions are frequently used have stability calculation and have a high degree of versatility, for example, matrix multiplication, and calculus. In addition, there are some commonly used methods not necessarily parallel methods, are added to the method library, which can effectively increase the development efficiency, such as the solution process of trigonometric, exponential and logarithmic functions, algebraic equations, differential equations and integral equations etc.

Special library provides a dedicated parallel interface. For different areas of research, application developers extract a special interface of methods such as the forward modeling of wave equation in the seismic exploration field, one-way wave equation pre-stack depth migration and analysis of ocean current fluctuations in ocean forecasting system, etc. Only after rigorous testing, the test method can be stored to the special library. The proven programs can not only ensure the correctness of the calculation results, but also respond to the error calling.

In the method library management module, we mainly complete function of the establishment, update, and search of the method, in addition to the internal communication of method library and management of the method dictionary. Its main function is to control the running of the method library, concluding the management sub-module of method library, the sub-module of method access, the sub-module of method update, the sub-module of the method link.

As noted in Fig. 29.2, this is an overall block diagram of the method library system. The method base system can be divided into Method Library and Method Library Base Management System.

In the Method Library part, there are two parts. First, Program Library, which is used to store the programs of algorithms and functions, including the program un-built and built. In the practical application, the programs are always built into the object files or the lib-files, so that it can save pretty long time when the functions are called. And then, the library dictionary is another necessary part, and its utility is store all the information of these functions or algorithms, such as the name of the functions, the parallel tools used, the utility and so on.

The second part is Method Library Management System. This sub-system communicates with the Library Dictionary, and executes its function, for example: retrieval the algorithms, update and modify. When the user add new functions or modify, the sub-system can communicate with the Library Dictionary, and update the information, and modify or delete the source codes and lib-files.

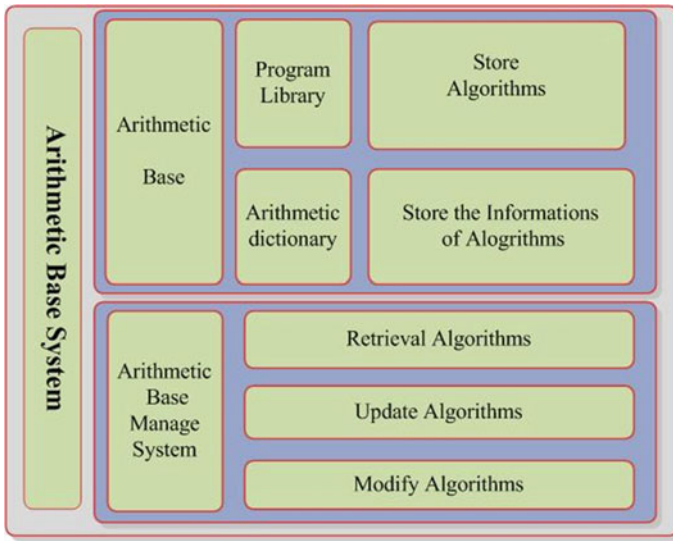


Fig. 29.2 Method library system structure chart

When using the parallel model based on method library, method library system can effectively provide the methods or functions in use, which can not only reduce development effort, but also makes the program easy to understand and reduce the probability of errors and effectively manage the internal methods.

### 29.4 Experiment

In the petroleum exploration industry, according to the tests of experimental data preprocessing module in the rapid modeling of seismic body, the program which is dealing with the coordinate pretreatment, we use the following samples: single-

Table 29.1 Comparison of cost time

File size (MB)	1-Node nothing (s)	4-Node only MPI (s)	4-Node all (s)
10	1.21	0.32	0.22
20	2.39	0.64	0.53
30	3.53	0.95	0.82
40	4.74	1.19	1.06
50	5.92	1.48	1.37
60	7.01	1.77	1.65
70	8.09	2.06	1.84
80	9.01	2.35	2.05
90	10.01	2.61	2.38
100	10.5	2.94	2.71



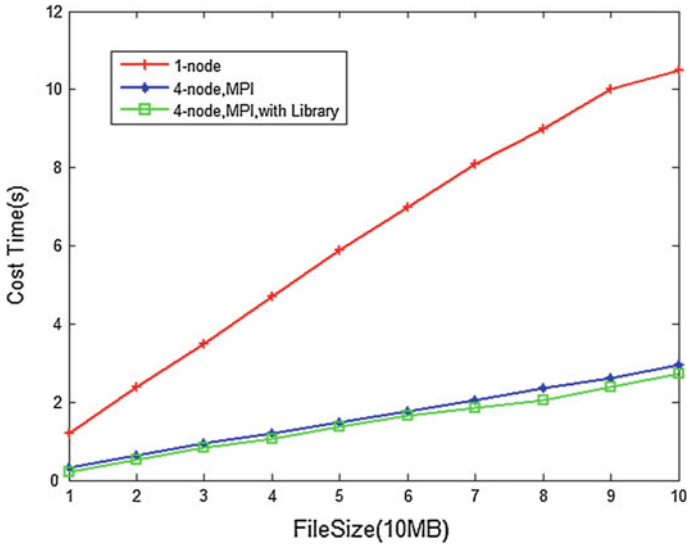


Fig. 29.3 Result of comparison-1

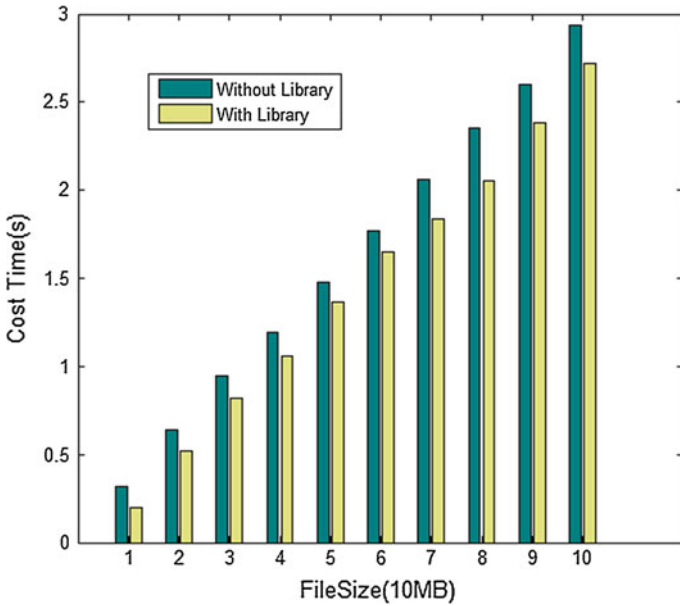


Fig. 29.4 Result of comparison-2

node non-parallel machine, four nodes MPI parallel machine, four nodes MPI parallel using method library. The results of cost time in different condition compared are showed in Table 29.1.

Figures 29.3 and 29.4 respectively show the result of the comparison with Matlab. Figure 29.3 is the comparison of the contrast in all three cases, and Fig. 29.4 is the comparison result of whether using the method library or not.

After the introduction of the method library, the speedup can be achieved on this issue, and we can calculate the average speedup is 1.16.

## 29.5 Conclusion

The obvious disadvantage of the parallel computing model of hierarchical heterogeneous is the usage of high-level language to write code is complex and error-prone. The idea of a method-based base system is put forward to modify the model and solve the problems. The method library based on hierarchical heterogeneous improves by adding the method library, offers a tool that can calculate, analysis and process; improve the computational efficiency of the process as a whole; achieve code reuse. This not only reduces the difficulty of program development, and more effective to speed up processing. So the method library application is extraordinary significant, in the field of both theory and practical application.

**Acknowledgments** This work is supported by Study of the Security Issues of Universal Calculate Trust model Based on Game Theory (Doctoral Fund of the Ministry of Education) under Grant 20100132110009.

## References

1. Chen G, Sun G, Xu Y, Long B (2009) Integrated parallel computing research status and development trend. *Chin Sci Bull* 54(8):1043–1049 (in Chinese)
2. Qiao XZ, Chen SQ, Yang LT (2004) A hierarchical model for parallel computations. *Int J High Perform Comput Networking* 1:117–127
3. Mattson TG, Sanders BA, Massingill BL (2004) *Patterns for parallel programming*. Addison Wesley Professional, Reading
4. Sun XH (2008) Scalable computing in the multi-core era. In: *Proceedings of the inaugural symposium on parallel algorithms, architectures and programming*. University of Science and Technology of China Press, Sep 2008, pp 1–18
5. Chen G, Sun G, Xu Y (2008) Parallel algorithms research methodology. *J Comput* 31(9):1493–1502 (in Chinese)
6. Chen G, Mao Q, Sun G, Xu Y (2008) Layered models of parallel computation. *J Univ Sci Technol China* 38:1–7 (June 2008) (in Chinese)
7. Ghemawat S, Gobioff H, Shun-Tak L (2003) The Google file system. In: *Proceedings of the nineteenth ACM symposium on operating systems principles*, New York. ACM Press, New York, 19–22 Oct 2003, pp 29–43

8. Zhu Z (2011) Compiler lexical analyzer to achieve. *Changchun Inst Technol (Nat Sci)* 24:132–134 (in Chinese)
9. Cui H, Dai G, Wang H, Zhang S (2004) Dynamic compiler technology. *Comput Sci* 31(7):113–117 (in Chinese)
10. Heinzelman WR, Chandrakasan A, Balakrishnan H (2000) Energy, efficient communication protocol for wireless micro-sensor networks. In: *Proceedings of the 33rd international conference on system sciences (HICSS'00)*, Jan 2000, pp 1–10

# Chapter 30

## Research on Weakly-Supervised Entity Relation Extraction of Specific Domain Based on Entropy Minimization

Jun Zhao, Jianyi Guo, Zhengtao Yu, Peng Chen and Cunli Mao

**Abstract** There are two major issues of automatic entity relation extraction: human intervention and difficulty in labeling corpus. For these two problems, combined with the characteristics of the tourism domain, this paper adopts a weakly-supervised extraction method of entity relation based on entropy minimization. This method firstly extracts the characteristic words by the idea of scalar clustering with small-scale stratified marked instances, and constructs the initial classifier with maximum entropy machine learning algorithm. Then use the initial classifier of certain accuracy to classify the unlabeled instances, and add the instances of the minimum information entropy to the training corpus set to continually expand the scale of training data set. Finally, repeat the above iterative process until the performance of classifier is to be stabilized, and then a final extraction classifier of entity relation in specific domain will be constructed. Experiments performed on the corpus of tourism domain show that, not only can this method reduce the dependence of entity relation extraction on manual intervention, but it could effectively improve the performance of entity relation extraction, the F value of which is up to 63.69 %.

**Keywords** Entity relation extraction · Entropy minimization · Weakly-supervised · Specific domain

### 30.1 Introduction

Entity relation extraction is not only an important part of information extraction, but extracted entity relation and event relation can be used to establish a more effective and compact information retrieval index so that improves the precision

---

J. Zhao · J. Guo (✉) · Z. Yu · P. Chen · C. Mao  
The School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650051, China  
e-mail: gjade86@hotmail.com

and speed of information retrieval. The supervised methods of semantic relation extraction in literature [1–6] have achieved a better extraction performance, but these methods depend on manual annotated corpus. The unsupervised methods of semantic relation extraction [7] don't need predefined relation category and the manual annotation data, but have a poor performance and can not identify the class of each instance in the learning process. To solve these problems, weakly-supervised learning methods are referenced to extract entity relation, which only need to provide a small amount of labeled instances as the initial training collection to get a better performance through effectively using large-scale unlabeled corpus. However, there is a gap between the performance of weakly-supervised methods and the supervised methods to semantic relation extraction. Therefore how to induce an extraction system of weakly-supervised entity relation of higher performance by a small amount of labeled instances has become a problem which is needed to study. Some scholars have proposed universal machine learning algorithms for entity relation extraction. Xibin et al. [8] proposed the weakly-supervised method of semantic relation extraction based on Bootstrapping, which proposed how to select instances of high credibility to start the next iteration, and proved the importance of the initial training set by performing experiments on ACE2004 marked corpus.

To solve the problems existed in above methods of how to select initial corpus and new instances of higher credibility, learning from literature [8–10], this paper, taken the domain of tourism as the background, adopts a weakly-supervised entity relation extraction method based on the minimum information entropy value (defined as entropy minimization in this article). The method makes full use of maximum entropy machine learning algorithm for flexible feature selection and strong portability model in the different domains, to build an initial maximum entropy classifier with certain accuracy by optimizing the context characteristics of each entity relation instance in relation extraction model. At the same time calculate the information entropy (i.e., confidence) of each instance to determine whether the instance is added to the training corpus, if the information entropy value of the instance is smaller, the certainty is greater, the credibility is higher, so continually adding these instances can automatically expand the scale and improve the quality of training data.

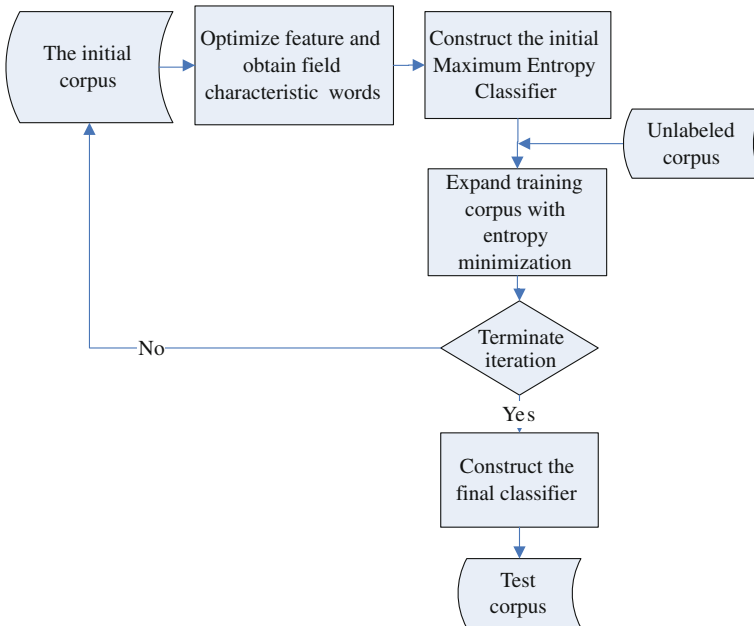
## **30.2 Weakly-Supervised Entity Relation Extraction of Specific Domain Based on Entropy Minimization**

Taken the domain of tourism as ground and combined with domain vocabulary, context, entity relation types, this paper takes the view of the machine learning algorithm modeling and expansion of the training data. The entity relation extraction process of tourism domain based on entropy minimization and weakly-supervised is described as follows. This article classifies entity relation types in

tourism domain four major categories, including the geographical relation, structural affiliation, property limited relation, the relation between festival and date. So artificially and stratifiedly select the proportional statements of marked instances, each kind of selection has 200 cases. Secondly, in the process of using the maximum entropy machine learning algorithm to construct the initial classifier, optimize the selection of eigenvectors that includes syntax, semantics and other characteristics. Calculate the semantic similarity of the seed words and vocabularies with the bootstrapping idea of “common neighbor” in scalar clustering to choose the optimal terms of large similarity as semantic words in specific domain. Thirdly, add the relation instance of the highest credibility to the initial training data set with entropy minimization method, thereby expanding the training data set. Finally, when the performance of maximum entropy classifier tends to be stabilized, terminate the iteration and at the same time obtain the final maximum entropy classifier. The entire process is shown in Fig. 30.1.

### 30.2.1 The Selection of Initial Data Set

Through in-depth analysis of the corpus of tourism domain, entity relation in the tourism domain could be classified. In this paper, divide the types of entity relation



**Fig. 30.1** The process of weakly-supervised entity relation extraction of specific domain based on entropy minimization

**Table 30.1** The definition of entity relation type in tourism domain

Entity relation major class	Entity relation subclass
Geographical relation	Contain, be adjacent to, be away from
Structural affiliation	Produce, belong to
Property limited relation	Ticket price, cover an area of, altitude, temperature, humidity, the best season, attractions star, hotel star, custom
The relation between festival and date	Holding time

into four major categories: geographical relation, structural affiliation, property limited relation, the relation between festival and date. Each relation type is defined, as shown in Table 30.1. For the selection of the initial training set, this paper uses the stratified strategy to select proportional corpus of four relation types, each of 200 cases, a total of 800 cases, in order to solve the data sparseness problem.

### 30.2.2 The Construction of Initial Maximum Entropy Classifier

How to get special semantic vocabularies combined the characteristics of tourism domain and improve the performance of relation recognition?

By the analysis of four major entity categories, we find that each type has its own special words which fully express the relation between two pairs of entity. These terms are listed below: geographical relation includes “be located”, “be adjacent to” and some other verbs. The structural affiliation includes “have”, “produce” and other verbs. The relation between festival and date includes “hold”, “celebrate” and other verbs. However property limited relation includes the vocabulary of “tickets”, “altitude”, “open time”, “the maximum temperature” and other nouns or compound nouns. Based on the above characteristics, this paper uses these vocabularies as seed template of four relation types, then uses scalar clustering algorithm to calculate the similarity of vocabulary before and after the entity and seed template in the sentence, and thus obtains more vocabularies with domain characteristics to be eigenvectors of relation classification.

$$\text{sim}(x, y) = \cos(x, y) = \frac{x \cdot y}{|x||y|} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (30.1)$$

Add the words of the maximum similarity to the feature set through calculating the similarity of many vocabularies and seed words with formula 30.1, until the training corpus set is empty. Finally, manage the above all selected eigenvectors to training format, and build the initial maximum entropy classifier.

### ***30.2.3 Training Corpus Expansion Based on Entropy Minimization and Termination of Iterations***

#### **30.2.3.1 Training Data Expansion Based on Entropy Minimization**

Information entropy of the random events:

$$H(\xi) = - \sum_1^n p_i \log p_i \quad (30.2)$$

Apply formula 30.2 of information entropy value in the entity relation extraction, and then the entropy value of each instance could be evaluated with formula 30.2. Where  $A_i$  represents the  $i$ th type of entity relation,  $p_i$  represents the probability of the  $i$ th class that the current instance is assigned to,  $n$  represents the number of semantic relation categories. According to formula 30.2, after each iteration calculate the entropy value  $H$  of each unlabeled instance to be classified. In accordance with the definition of entropy, the smaller the entropy, the better the certainty, the higher the instance classification credibility, so the possibilities of correct classification are higher. Add the above instances to the training set to the next round iteration, and then the risk will be smaller.

#### **30.2.3.2 The Maximum Entropy Classifier Construction of Entity Relation Extraction**

The iterative process of maximum entropy classifier construction of entity relation extraction: in accordance with entropy minimization method, add new labeled instances obtained each time to training data set with the initial maximum entropy classifier. Repeat the process up to meet the iteration termination condition, that is to say the performance of the classifier is to be stabilized. Now the final classification of entity relation extraction in specific domain has been constructed.

## **30.3 Experimental Design and Results**

### ***30.3.1 Experimental Design***

This article uses annotated corpus collected by ourselves in tourism domain as experimental data, which are collected by Web crawler tool from the network. The paper chooses named entities and instances of apparent semantic relations in tourism domain as the experimental data. Here are experimental steps: First remove html tags, segment sentences and words, tag POS, then remove stop words and other pretreatments to format the preliminary experimental data. Second,



stratifiedly select proportional instances of the four relation types, each of type has 200 cases. Third, recognize named entities from the above corpus, and calculate semantic similarity of vocabularies in the seed mode and unlabeled instances with the scalar clustering method, then make the vocabularies of the highest semantic similarity and syntax and other characteristics as eigenvectors of building a classifier, and process them to the data format needed by the experiment. Finally, classify the unlabeled instances with the above initial classifier, and the instance of entropy minimization is added to the training set after calculating the information entropy of each instance, however the remaining unlabeled instances will wait for the next classification, then repeat the above iteration process until the classifier performance is in a steady state, by then the final classifier will be built.

This paper uses Maximum Entropy Modeling Toolkit for Python and C++<sup>1</sup> developed by Dr. Zhang Le.

### ***30.3.2 Experimental Results and Analysis***

#### **30.3.2.1 The Impact of Specific Vocabulary Eigenvectors**

In this paper, the characteristic vocabulary calculated by the idea of scalar clustering is one of the eigenvectors constructing classifier. Experiment 1 is the experiment of relation extraction classifier construction with no characteristic words. Experiment 2 is the experiment of relation extraction classifier construction with characteristic words. Experimental comparisons are shown as Table 30.2. From Table 30.2, we can conclude that semantic vocabulary as the characteristics of building a classifier can significantly improve the extraction performance.

#### **30.3.2.2 The Comparisons with Different Information Entropy**

In this paper, the unlabeled instances of the minimum information entropy as the highest credibility instances are added to the training data set, in order to ensure expand the scale of the training data set on the basis of better classifier performance. Experiment 4 shows the comparison between the method this paper adopts and other experimental methods (information gain method, SVM and information entropy method), as shown in Table 30.3.

Table 30.3 shows that, compared with the methods in the literature 9, under the same performance of initial maximum entropy classifier, the effect of entity relation extraction with entropy minimization method is better than the information gain method to expand the training data. Compared with the literature 10, select the corpus according to the stratified strategy, then the performance of extraction classifier is better.

---

<sup>1</sup> [http://www.homepages.inf.edu.ac.uk/lzhang10/maxent\\_toolkit.html](http://www.homepages.inf.edu.ac.uk/lzhang10/maxent_toolkit.html)

**Table 30.2** The experimental results of semantic vocabulary to entity relation extraction in tourism domain

Experimental methods	Entity relation class	The number of correct classification	Predict the total number	The total number in test corpus	Precision/%	Recall/%	F value/%
Experiment 1/	No relation	1458/1498	1532/1546	1588/1588	91.83/94.31	95.19/96.89	93.48/95.58
Experiment 2	Geographical relation	521/535	765/763	731/731	71.30/73.12	68.13/70.01	69.68/71.53
	Structural affiliation	369/387	581/555	551/551	66.90/70.18	63.42/69.63	65.11/69.90
	Property limited relation	197/210	293/322	316/316	62.19/66.47	67.01/65.22	64.51/65.84
	The relation between festival and date	253/251	343/312	359/359	70.53/69.85	73.92/80.50	72.19/74.80

**Table 30.3** The experimental comparison results of relation extraction with different ways to expand the training data

Experimental method	Entity relation class	Predict the total number	The number of correct classification	The total number in test corpus	Precision/ %	Recall/ %	F-Score/ %
Information gain method/ SVM and information	Geographical relation	316/321/352	150/174/191	252/252/278	47.37/54.13/ 54.26	59.38/69.02/ 68.71	52.70/60.67/ 60.64
entropy method/ Minimum entropy method	Structural affiliation	828/989/881	501/578/601	952/952/984	60.51/58.46/ 68.22	52.63/60.74/ 61.08	56.30/59.58/ 64.45
	Property limited relation	218/231/301	134/155/195	278/278/290	61.47/67.07/ 64.78	48.20/55.79/ 67.24	54.03/60.91/ 65.98

## 30.4 Conclusion

In this paper, by means of information entropy theory, first entity relation extraction classifier was constructed with maximum entropy model, then, according to the information entropy value of each instance belonged to the class, determined whether the instance was added to the training corpus. The practices have proved that this method can efficiently extract entity relation from the texts of specific domains. The next step is combining with domain characteristics to explore a better method of weakly-supervised entity relation extraction, and further improve the performance of relation extraction.

**Acknowledgments** This paper is supported by National Nature Science Foundation (No. 60863011, 61175068), and the Key Project of Yunnan Nature Science Foundation (No. 2008CC023), and the National Innovation Fund for Technology based Firms (No. 11C26215305905), and the Open Fund of Software Engineering Key Laboratory of Yunnan Province (No. 2011SE14), and the Ministry of Education of Returned Overseas Students to Start Research and Fund Projects.

## References

1. Kambhatla N (2004) Combining lexical, syntactic and semantic features with maximum entropy models for extracting relations. In: Proceedings of the ACL 2004 on interactive poster and demonstration sessions, pp 178–181
2. Zhou GD, Su J, Zhang J (2005) Exploring various knowledge in relation extraction. In: Proceedings of the 43rd annual meeting on association for computational linguistics, pp 427–434
3. Zhao SB, Grishman R (2005) Extracting relations with integrated information using kernel methods. In: Proceedings of the 43rd annual meeting on association for computational linguistics, pp 419–426
4. Wang T, Li YY, Kalina B (2006). Automatic extraction of hierarchical relations from text. In: Proceedings of the third European semantic web conference, pp 401–416
5. Che WX, Liu T, Li S (2005) Automatic entity relation extraction. *J Chin Inf Process* 19(2):1–6 (in chinese)
6. Dong J, Sun L, Feng YY, Huang RH (2007) Chinese automatic entity relation extraction. *J Chin Inf Process* 21(4):80–85,91
7. Huang C, Qian LH, Zhou GD, Zhu QM (2010) Research on unsupervised Chinese entity relation extraction based on convolution tree kernel. *J Chin Inf Process* 24(4):11–17
8. Xi B, Zhou GD, Qian LH, Pan S (2008) Weakly-supervised semantic relation extraction using stratified strategy. *J Guangxi Norm Univ* 26(1):178–181 (in Chinese)
9. He TT, Xu C, Li J (2006) Named entity relation extraction method based on seed self-expansion. *Comput Eng* 32(21):183–184 (in Chinese)
10. Zhang Z (2004) Weakly-supervised relation classification for information extraction. In: Proceedings of the 13th conference on information and knowledge management (CIKM'2004), pp 581–588

# Chapter 31

## Using Fast Sampling-Insensitive Stereo Matching For 2-D Face Recognition

Rui Liu, Longfei Cui, Wenke Zhang and Ming Zhu

**Abstract** In this paper, we propose using sampling-insensitive stereo matching for 2-D face recognition. We don't perform 3-D reconstruction but define a measure of the similarity of two faces. Then we match one 2-D query image to one 2-D gallery image using the measure for face recognition. We show that this method is not only robust to pose variations but also faster than other stereo matching methods. The proposed approach has been tested on the CMU PIE and ORL data set and demonstrates superior performance compared to existing methods in real-world situations including changes in pose and illumination.

**Keywords** Face recognition · Sampling-insensitive · Stereo matching · Fast

### 31.1 Introduction

Face recognition has attracted much attention due to its potential value for applications and its theoretical challenges. Although face recognition in controlled environment has been well solved, its performance in real application is still far from satisfactory. The variations of pose, occlusion, expression and illumination are still critical issues that affect the face recognition performance.

It is common to apply existing techniques such as Eigenfaces [1] or Fisherfaces [2] to reduce the dimension of the face images. However, these techniques are not robust to outliers or extreme noise such as pose and illumination variations [3]. Local features such as Gabor [4] and local binary patterns (LBP) [5] are more stable to local changes. Some recent work mainly focused on the improvement of

---

R. Liu (✉) · L. Cui · W. Zhang · M. Zhu  
Department of Electronic Engineering and Information Science, University of Science and Technology of China, RM 10-627, USTC, P.O. Box 4, Hefei 230027 Anhui, People Republic of China  
e-mail: liuruin@mail.ustc.edu.cn

the discriminative ability of LBP like features [6]. But these methods are not initially designed to handle variations in pose.

Castillo and Jacobs [7] have shown that stereo matching algorithms can gain significant performance when used to perform 2D face recognition in the presence of pose variation. However, this algorithm works at pixel resolution and still suffers from sampling effects [8]. It also degrades gracefully with changes of light [7].

In this paper, we propose a measure of the similarity of two faces (in unknown poses) that is not only fast but also insensitive to image sampling. We show that the matching cost is robust to horizontal pose variations and illumination variations. We also show that the method works well even when the epipolar lines we use are not accurate.

The paper is organized as follows. We introduce the dissimilarity measure in pixel level and describe its computation in Sect. 31.2. The measure is incorporated into a stereo algorithm for face recognition in Sect. 31.3. Section 31.4 presents and analyses all experiments. Finally, in Sect. 31.5, conclusions will be given.

## 31.2 Sampling-Insensitive Dissimilarity Measure

Most work in image-based recognition aligns regions to be matched with a low-dimensional transformation, such as translation, or a similarity or affine transformation. Instead, we use stereo matching.

We require an efficient stereo algorithm appropriate for wide baseline matching of faces. Castillo and Jacobs [7] have used the four planes, four transitions dynamic programming stereo matching algorithm as proposed by Criminisi et al. [9]. We don't do this because of the sampling effects and the light influence. Instead, we use Birchfield et al. [8] which has been developed for improving the traditional measures of pixel dissimilarity in stereo matching. It is not obvious that it will work for robust face recognition, but we will show that it does.

The core of this method calculates a measure of pixel dissimilarity that compares two pixels using the linearly interpolated intensity functions surrounding them. The method defines the dissimilarity  $d$  between the pixels as the following:

$$d(x_L, x_R) = \min\{\bar{d}(x_L, x_R, I_L, I_R), \bar{d}(x_R, x_L, I_R, I_L)\} \quad (31.1)$$

Here,  $\bar{d}(x_L, x_R, I_L, I_R)$  is used to measure how well the intensity at position  $x_L$  in the left scanline fits into the linearly interpolated region surrounding at position  $x_R$  in the right scanline:

$$\bar{d}(x_L, x_R, I_L, I_R) = \min_{x_R - \frac{1}{2} \leq x \leq x_R + \frac{1}{2}} |I_L(x_L) - \hat{I}_R(x)| \quad (31.2)$$

$I_L$  and  $I_R$  are two discrete one-dimensional arrays of intensity values sampled by the ideal sampler of the image sensor.  $\hat{I}_R$  is the linearly interpolated function

between the sample points of the right scanline, and  $\hat{I}_L$  is defined similarly. So we obtain a symmetric quantity:

$$\bar{d}(x_R, x_L, I_R, I_L) = \min_{x_L - \frac{1}{2} \leq x \leq x_L + \frac{1}{2}} |\hat{I}_L(x) - I_R(x_R)| \quad (31.3)$$

Since the extreme points of a piecewise linear function must be its breakpoints, the computation of  $d$  is straightforward. First, we compute the linearly interpolated intensity halfway between  $x_R$  and its neighboring pixel to the left:

$$I_R^- \equiv \hat{I}_R\left(x_R - \frac{1}{2}\right) = \frac{1}{2}(I_R(x_R) + I_R(x_R - 1)) \quad (31.4)$$

And the analogous quantity:

$$I_R^+ \equiv \hat{I}_R\left(x_R + \frac{1}{2}\right) = \frac{1}{2}(I_R(x_R) + I_R(x_R + 1)) \quad (31.5)$$

Then, let  $I_{\min} = \min\{I_R^-, I_R^+, I_R(x_R)\}$  and  $I_{\max} = \max\{I_R^-, I_R^+, I_R(x_R)\}$ . With these quantities defined,

$$\bar{d}(x_L, x_R, I_L, I_R) = \max\{0, I_L(x_L) - I_{\max}, I_{\min} - I_L(x_L)\} \quad (31.6)$$

And it has been proved in [8] that this computation, along with its symmetric counterpart  $\bar{d}(x_R, x_L, I_R, I_L)$ , takes only a small and constant amount of time more than the absolute difference in intensity.

As in [7], it is important to stress that we are relatively unaffected by some of the difficulties that make it hard to avoid artifacts in stereo reconstruction. Selecting the right match is difficult, but important for good reconstructions. However, since we only use the cost of a matching, selecting the right matching is unimportant to us in this case. Also, errors in small regions, such as at occluding boundaries, can produce bad artifacts in reconstructions, but that is not a problem for our method as long as they don't affect the cost too much.

### 31.3 Stereo Matching for Face Recognition

Though we have got the measure of the dissimilarity between the pixels, it is still far from face recognition. We need to modify it and incorporate it into a stereo algorithm so that we can calculate a similarity between two face images.

First, we extend the notation  $d(x_L, x_R)$  in (31.1) to  $d(\langle x_L, y_L \rangle, \langle x_R, y_R \rangle)$ , where  $y_L$  and  $y_R$  are the scan line (row) of left image and right image respectively. So we can take  $d(x_L, x_R)$  as the simplified form of  $d(\langle x_L, y_L \rangle, \langle x_R, y_R \rangle)$  where  $y_L = y_R$ .

Let  $I_1(x, y)$  and  $I_2(x, y)$  be a given pair of rectified stereo face images. The absolute intensity difference image is found using Eq. (31.7), where  $\alpha$  denotes the

relative horizontal shift between the two input faces,  $\beta$  denotes the relative vertical shift between the two input faces, and  $\omega$  is the experimental parameter to adjust the distribution.

$$\Delta I(x, y, \alpha, \beta) = e^{-\omega * d(\langle x+\alpha, y+\beta \rangle, \langle x, y \rangle)} \quad (31.7)$$

Second, given the set of possible horizontal shift A and the set of possible horizontal shift B, we compute a matching value  $m(I_1, I_2)$  which tells us how well image  $I_1$  and image  $I_2$  match:

$$m(I_1, I_2) = \frac{1}{w * h} \sum_{x=0}^w \sum_{y=0}^h \max_{\alpha \in A} \max_{\beta \in B} [\Delta I(x, y, \alpha, \beta)] \quad (31.8)$$

where  $w$  is the width of the image and  $h$  is the height of the image.

Third, given two images  $I_1$  and  $I_2$ , we define the similarity of the two images as:

$$\text{similarity}(I_1, I_2) = \max \begin{cases} m(I_1, I_2) \\ m(I_2, I_1) \\ m(\text{flip}(I_1), I_2) \\ m(I_2, \text{flip}(I_1)) \end{cases} \quad (31.9)$$

Since we do not know which image is left and which image is right we have to try both options, one of them will be the true cost, and the other cost will be noise and should be ignored. Here, flip produces a left–right reflection of the image.

It has been showed that flip is helpful when two views see mainly different sides of the face [7]. In this case, a truly correct correspondence would mark most of the face as occluded. However, since faces are approximately vertically symmetric, flip approximates a rotation about the y axis that creates a virtual view so that the same side of the face is visible in both images. For example, if we viewed a face in left and right profile, there would be no points on the face visible in both images, but flipping one image would still allow us to produce a good match.

Finally, we perform recognition simply by matching a probe image to the most similar image in the gallery.

## 31.4 Experiments

### 31.4.1 CMU PIE Database

The CMU PIE [10] database consists of 13 poses of which 9 have approximately the same camera altitude (poses: c34, c14, c11, c29, c27, c05, c37, c25 and c22). Three other poses that have a significantly higher camera altitude (poses: c31, c09 and c02) and there is one last pose that has a significantly lower camera altitude (pose c07).



We conducted an experiment to compare our method with five others. We compared with two variants of eigen light-fields [11], eigenfaces [1], dynamic programming (DP) based stereo matching [7] and FaceIt as described in [11, 12]. FaceIt (Version 2.5.0.17, as in [7]) is a commercial face recognition system from Identix which finished top overall in the Face Recognition Vendor Test 2000. Eigenfaces is a common benchmark algorithm for face recognition. Eigen light-fields is a state of the art method for face recognition across pose variation. Finally, dynamic programming based stereo matching (DP-SM) is a traditional stereo matching method. But it is used in face recognition for the first time and shows, to our knowledge, the best reported results on a subset of 34 people of the PIE database for pose variation.

We set the parameters as  $\omega = 20$ ,  $A = [0, 5]$ ,  $B = [0, 5]$ . In order to compare the methods in equal condition, we only tested on individuals 35–68 from the PIE database. Specifically, we selected each gallery pose as one of the 13 PIE poses and the probe pose as one of the remaining 12 poses, for a total of 156 gallery-probe pairs. We evaluated the accuracy of our method in this setting and compared to the results in [7, 11, 12]. Table 31.1 summarizes the average recognition rates.

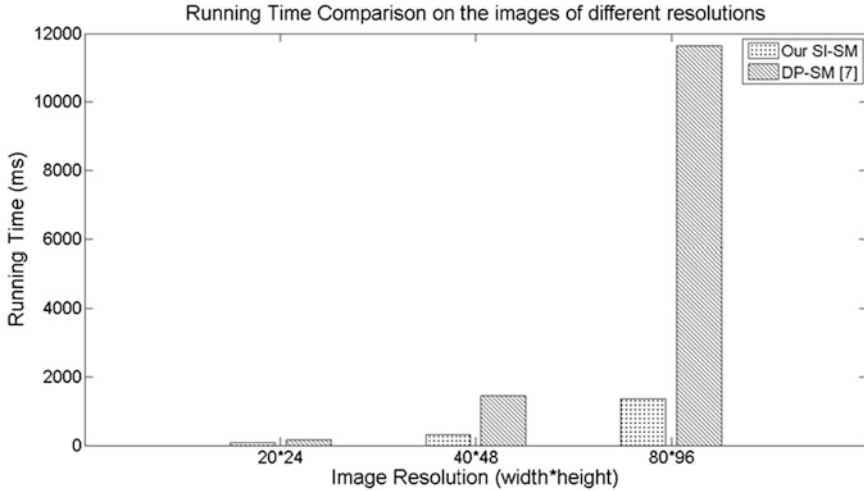
When compared with [7], our SI-SM method shows 1.7 % losses. However, experiments on the images of different resolutions demonstrate that our method is much faster than DP-SM (the source code was provided by Castillo). Details can be seen in Fig. 31.1. We conducted these experiments on Intel(R) Core(TM) i5-2467 M CPU(1.60Ghz) and 4 GB DDR3 RAM. The two methods were both implemented using C/C++. Notice that the running time of SI-SM is linearly growing with the growth of resolution while DP-SM is not.

### 31.4.2 *ORL Database*

We also evaluate our method on Olivetti Research Laboratory (ORL) face image database. This ORL face database acts as the reference face database and contains 400 face images. There are 10 face images each of 40 different persons. For some of the subjects, the images were taken at different times, varying lighting slightly, facial expressions (open/closed eyes, smiling/non-smiling) and facial details (glasses/no-glasses).

**Table 31.1** A comparison of our Sampling-Insensitive Stereo Matching (SI-SM) approach with other methods on 34 faces of PIE database

Method	Accuracy (%)
Eigenfaces [11, 12]	16.6
FaceIt [11, 12]	24.3
Eigen light-fields (3-point norm.) [11, 12]	52.5
Eigen light-fields (Multi-point norm.) [11, 12]	66.3
DP-SM [7]	82.0
<b>Our SI-SM</b>	<b>80.3</b>



**Fig. 31.1** The running time of face recognition using SI-SM and DP-SM was measured on i5-2467 M CPU (1.60Ghz) and 4 GB DDR3 RAM. For each resolution, eight pairs of images were picked out. We computed the similarity of each probe-gallery pair and recorded the time. The final running time for each resolution is then obtained by calculating the average

All the images are taken against a dark homogeneous background and the subjects are in upright, frontal position with tolerance for some tilting and rotation of up to  $20^\circ$ . Moreover, the most variation of some image scale is close to 10 %. The size of each image is  $92 \times 112$ , 8-bit grey levels.

We compare our proposed method against well-established existing techniques as presented in [13] like standard eigenface based methods [14], wavelet [15], curvelet based methods [13], and discriminant wavelet-faces using nearest feature line (NFL) classifier [16]. LBP [5] method is also compared to our algorithm.

To be able to compare results with these methods as presented in [5], we needed to use a subset of 200 images because they randomly chose five face images per person for training and the remaining five images for testing. We don't require training, but we're interested in comparing the methods in equal condition so we randomly tested on 200 images (five face images per person) from the ORL database. All the faces are then scaled to the size  $128 \times 128$  pixels, aligned according to the eye positions.

The results in Table 31.2 demonstrate that high recognition accuracy can be achieved using our SISM method for 2-D face recognition. Our technique has been found to be robust against variation in pose and illumination as well. When compared to the best performing method in table (Curveletface + PCA + LDA), our SI-SM method shows slight losses. However, our method is much simpler and more straightforward, which is very important for applications.

**Table 31.2** A comparison of our Sampling-Insensitive Stereo Matching (SI-SM) approach with other methods on ORL database

Method	Accuracy (%)
Standard eigenface [14]	92.2
LBP [5]	97.5
Waveletface [15]	92.5
Curveletface [13]	94.5
Waveletface + PCA [15]	94.5
Waveletface + LDA [16]	94.7
Waveletface + LDA + NFL [16]	95.2
Curveletface + PCA [13]	96.6
Curveletface + PCA + LDA [13]	97.7
<b>Our SI-SM</b>	<b>97.0</b>

## 31.5 Conclusion

We proposed a method for 2-D face recognition. Compared to existing methods, ours is very simple and performs very well. Though we also used stereo matching for recognition and obtained about the same accuracy as [7], our method has been proved to be much faster. There is still a lot of room for improvement in our method. For example, it remains a future direction to determine how best to incorporate learning into it. Some strategies can also be pursued to deal with illumination variations.

**Acknowledgments** This research was supported by the “Strategic Priority Research Program-Network Video Communication and Control” of the Chinese Academy of Sciences (Grant No. XDA06030900), and by the National Key Technology R&D Program of China (Grant No. 2011BAH11B01).

## References

1. Turk M, Pentland A (1991) Face recognition using Eigenfaces. In: CVPR
2. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell* 19(7):711–720
3. De la Torre F, Black M (2001) Robust principal component analysis for computer vision. In: ICCV
4. Lei Z, Li SZ, Chu R, Zhu X (2007) Face recognition with local gabor textons. In: ICB
5. Ahonen T, Hadid A, Pietikäinen M (2004) Face recognition with local binary patterns. In: ECCV
6. Lei Z, Li SZ (2012) Learning discriminant face descriptor for face recognition. In: ACCV
7. Castillo CD, Jacobs (2007) DW using stereo matching for 2D face recognition across pose. In: CVPR
8. Birchfield S, Tomasi C (1998) A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Trans Pattern Anal Mach Intell* 20(4):401–406
9. Criminisi A, Blake A, Rother C, Shotton J, Torr PH (2007) Efficient dense stereo with occlusions for new view-synthesis by four-state dynamic programming. *Int J Comput Vis* 71(1):89–110

10. Sim T, Baker S, Bsat M (2003) The CMU pose, illumination, and expression database. *IEEE Trans Pattern Anal Mach Intell* 25(12):1615–1618
11. Gross R, Baker S, Matthews I, Kanade T (2004) Face recognition across pose and illumination. In: Li SZ, Jain AK (eds) *Handbook of face recognition*. Springer, New York
12. Gross R, Matthews I, Baker S (2004) Appearance based face recognition and light-fields. *IEEE Trans Pattern Anal Mach Intell* 26(4):449–465
13. Mandal T, Wu QMJ, Yuan Y (2009) Curvelet based face recognition via dimension reduction. *Signal Process* 89(12):2345–2353
14. Turk M, Pentland A (1991) Eigenfaces for recognition. *J Cogn Neurosci* 3(1):71–86
15. Feng GC, Yuen PC, Dai DQ (2000) Human face recognition using PCA on wavelet subband. *J Electron Imaging* 9(2):226–233
16. Chien JT, Wu CC (2002) Discriminant wavelet faces and nearest feature classifiers for face recognition. *IEEE Trans Pattern Anal Mach Intell* 24(2):1644–1649

# Chapter 32

## The Detection Method of Printed Registration Deviations Based on Machine Vision

Kailong Liu, Minrui Fei, Wenju Zhou and Haikuan Wang

**Abstract** In order to meet the actual requirements of judging four-color printed registration deviations quickly and accurately, this article describes a new detection method based on machine vision and the method has been designed involving three processes: removing interferential image by a corresponding region character separation (RCS) algorithm; detecting the edge of registration mark by interpolation subpixel algorithm and use weighted markov chain to calibrate the detection. The experiment indicates that the speed and accuracy with this method have greatly improved, and even with noise interference, this method can detect deviation quickly and accurately, superior to the traditional detection method.

**Keywords** Registration deviation · Machine vision · Region character separation · Interpolation subpixel algorithm · Weighted markov chain

### 32.1 Introduction

With the development of the printing technical level, the main direction of modern printing is to be more efficient, higher precision and higher quality [1]. In this process, detecting printed registration deviations quickly and accurately plays an important role in improving the automatic level of printing press [2, 3]. Using machine vision technology to judge the deviation has loads of advantages such as non-contact, high efficiency and high quality [4]. But in the high speed presswork

---

K. Liu (✉) · M. Fei · W. Zhou · H. Wang  
School of Mechanical Engineering and Automation, Shanghai University,  
No 140 GuangYan Road, ZhaBei, 200072 Shanghai City, China  
e-mail: 13701771896@163.com

M. Fei  
e-mail: mrfei888@x263.net

manufacture, the traditional detection method has great limitations and brings lots of unfavorable factors for presswork quality control [5]. Firstly, the mark with loads of interferential images may need complex algorithm to calculate and spend large amounts of time. Secondly, it is difficult to detect the edge of registration mark quickly and accurately when use the traditional algorithm such as least square (LS) algorithm, spatial moment (SM) algorithm etc. Thirdly, the results of detection may not the same as the real value of detection so it is necessary to use some statistics methods to decrease the error between the results and the real value.

This paper presents an intelligent detection method especially for four-color high speed presswork, and the method has been designed involving three processes: removing interferential image; detecting the edge of registration mark and use weighted markov chain to calibrate the detection. The paper is constructed as follows: the detection method is introduced particular in Sect. 32.2. In Sect. 32.3, weighted markov chain to calibrate the detection is introduced. Section 32.4 is the experiment test and results.

## 32.2 Details of the Detection Method

### 32.2.1 Four-Color Presswork Registration Mark Design

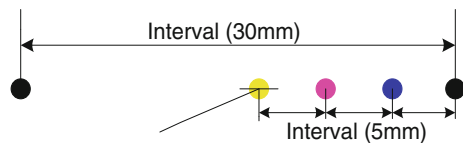
A kind of cross line registration mark was often used to calculate the deviation of presswork. But the cross line may overlap together after printing and is complex for the visual system to analyze. So a solid circle registration mark is designed in this paper to meet the actual requirements of speed and accuracy. The designed mark is shown in Fig. 32.1 and it includes two black solid circle, one yellow solid circle, one red solid circle and one blue solid circle (Table 32.1).

### 32.2.2 Remove Interferential Image

The interferential image is printed by some factors such as shaking of the machine, distorting of the printed matter and splashing in the process of printing registration mark pattern. A complete interferential image case is shown in Fig. 32.2.

In order to calculate the deviation of registration mark more accurately. A corresponding region character separation (RCS) algorithm is used to remove the interferential image. The process of removing black interferential images is divided into nine Steps as follows.

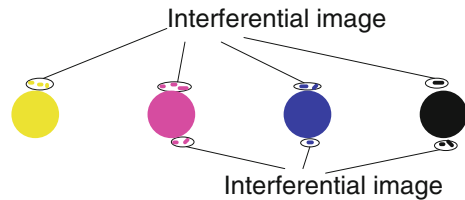
**Fig. 32.1** Registration marks schematic



**Table 32.1** The values of test distance among 21 times

Times	Distance (mm)	Level	Times	Distance (mm)	Level	Times	Distance (mm)	Level
1	5.017	4	8	5.010	2	15	5.008	2
2	5.008	2	9	5.021	5	16	5.004	1
3	5.012	3	10	5.016	4	17	5.003	1
4	5.014	3	11	5.009	2	18	5.006	2
5	5.012	3	12	5.008	2	19	5.013	3
6	5.012	3	13	5.016	4	20	5.017	4
7	5.008	2	14	5.010	2	21	5.011	3

**Fig. 32.2** Interferential image schematic



**Step 1.** Scanning each pixel of the image from left to right and from up to down.  
**Step 2.** Judging the scanning pixel. If it is a black spot, then executing the Step 3, or jumping back to Step 1.

**Step 3.** Judging the connectivity of spots in the area of upper-right, upper, upper-left to the current pixel according to the priority. Among this, the upper-right priority level is highest while the upper-left priority level is lowest.

**Step 4.** If values of upper-right spot and left spot (or upper-left point) are not the same, set the current spot the same value with the upper-right point, and then scanning images from the beginning to the end, putting the pixel whose value is the same as left spot (or upper-left spot) to the same value of upper-right spot.

**Step 5.** If the current pixel spot is black, and does not belong to Step 4, then set the current spot the same mark as well as the same value with the upper spot.

**Step 6.** If the left-upper spot is black, and does not belong to the Step 5, then set the current spot the same mark with the upper-left spot as well as the same value with left spot (Table 32.2).

**Step 7.** If the left spot is black, and does not belong to the Step 6, then set the current spot the same mark as well as the same value with the left spot.

**Step 8.** If the upper-right, upper, upper-left and left spot are not belong to black spots, then can recognize this spot is a new area point, then set current spot into the black spot, namely RGB value is (0, 0, 0).

**Step 9.** Remove the interferential image by a threshold value.

### 32.2.3 Detect the Edge of Registration Mark

The essence of the edge detection is to extract image gray discontinuous edge pixels through some algorithms [6]. In order to improve the accuracy of edge

detection, many experts put forward some effective algorithm such as least square (LS) algorithm, spatial moment (SM) algorithm and interpolation algorithm to make the image edge orientate to the subpixel level [7]. Least square (LS) algorithm and spatial moment (SM) algorithm have higher positioning accuracy, but longer computation time. Interpolation algorithm is relatively short time and the improvement of spline interpolation method, can effectively improve the detection speed.

The corresponding interpolation subpixel algorithm is shown in formula (32.1), and gains nearly on the best interpolation function.

$$S(w) = \begin{cases} 1 - 4|w|^2 + |w|^3 & |w| < 1 \\ -9|w| + 7|w|^2 + |w|^3 & 1 \leq |w| \leq 3 \\ 0 & |w| \geq 3 \end{cases} \quad (32.1)$$

$w$  is the spline node and the matrix represents two dimensional spline interpolation defined as

$$F(m, n) = ABC \quad (32.2)$$

where

$$A = \begin{bmatrix} s(1 + v) \\ s(v) \\ s(1 - v) \\ s(2 - v) \end{bmatrix}^T, \quad C = \begin{bmatrix} s(1 + u) \\ s(u) \\ s(1 - u) \\ s(2 - u) \end{bmatrix}$$

and

$$B = \begin{bmatrix} f(i - 1, j - 1) & f(i - 1, j) & f(i - 1, j + 1) & f(i - 1, j + 2) \\ f(i, j - 1) & f(i, j) & f(i, j + 1) & f(i, j + 2) \\ f(i + 1, j - 1) & f(i + 1, j) & f(i + 1, j + 1) & f(i + 1, j + 2) \\ f(i + 2, j - 1) & f(i + 2, j) & f(i + 2, j + 1) & f(i + 2, j + 2) \end{bmatrix}$$

$F(m,n)$  is the image after interpolating,  $f(i,j)$  is the imported pixel spot before interpolating, and  $u = m - [m]$ ,  $v = n - [n]$  respectively.

The clear continuous edge image can be picked up through the improvement spline interpolation subpixel algorithm and the level of edge image positioning accuracy is approximately 0.3.

### 32.3 Weighted Markov Chain to Calibrate the Detection

Registration deviation results of detection may not the same as the real value of detection and it may affect the servo mechanism to make accurate adjustment. A statistics method based on weighted markov chain is used in this paper to



calibrate the detection and decrease the error between the results and the real value.

Markov chain [8] is defined as  $\{X(n), n = 0, 1, 2, \dots\}$  while it should satisfy the formula (32.3) as follow.

$$\begin{aligned}
 P\{X(n_m + k) = j | X(n_1) = i_1, X(n_2) = i_2, \dots, X(n_m) = i_m\} \\
 = P\{X(n_m + k) = j | X(n_m) = i_m\}
 \end{aligned}
 \tag{32.3}$$

$P\{X(n + k) = j | X(n) = i\}, k \geq 1$  is called transition probability of  $K$  steps.

A normal distance between center of yellow circle and center of red circle should be 5 mm accurately. But in fact the results of detection may not the same as the real value of detection due to some factor such as less printed, shaking of the machine etc. So a experiment on a normal distance between yellow center and red center is tested for 21 times in this paper.

**Step 1.** Results of test distance is divided into 5 levels.

Level 1: 5.001–5.005 mm; Level 2: 5.006–5.010 mm; Level 3: 5.011–5.015 mm; Level 4: 5.016–5.020 mm; Level 5: 5.021–5.025 mm;

**Step 2.** Transfer frequency matrix (TFM)  $f_{ij}$  of the test distance and the transition probability matrix (TPM)  $P_{ij}$  are calculated as follow (Table 32.1).

$$f_{ij} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 3 & 2 & 1 & 1 \\ 0 & 1 & 3 & 1 & 0 \\ 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad P_{i,j(1)} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0.125 & 0.375 & 0.25 & 0.125 & 0.125 \\ 0 & 0.2 & 0.6 & 0.2 & 0 \\ 0 & 0.75 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

**Step 3.** Calculate the autocorrelation coefficient  $\rho_s$  to get the weight of every step. The formula is shown as.

$$\rho_s = \frac{\sum_{l=1}^{n-s} (x_l - \bar{x})(x_{l+s} - \bar{x})}{\sum_{l=1}^n (x_l - \bar{x})^2}
 \tag{32.4}$$

where  $x_l$  stands for the value of test distance,  $\bar{x}$  is the average value and the formula (32.4) is used to normalize the autocorrelation coefficient  $\rho_s$ ; Finally use formula (32.5) to sum up the weighted value as the final probability.

$$p_k = \left[ \sum_{s=1}^k p_k^{(s)}(1), \sum_{s=1}^k p_k^{(s)}(2), \dots, \sum_{s=1}^k p_k^{(s)}(k) \right]
 \tag{32.5}$$

**Step 4.** According to  $f_{ij}$  and  $P_{ij}$  calculated in Step 1, the statistical magnitude  $\chi^2$  could be calculated in formula (32.6).

$$\chi^2 = 2 \sum_{i=1}^5 \sum_{j=1}^5 f_{ij} \left| \ln \frac{P_{ij}}{P_j} \right| \tag{32.6}$$

**Step 5.** The transition probability matrix(TPM) between 2–5 are calculated as follows.

$$\begin{aligned}
 P_{ij2} &= \begin{bmatrix} 0.3125 & 0.4375 & 0.1250 & 0.0625 & 0.0625 \\ 0.1093 & 0.3469 & 0.2750 & 0.2219 & 0.0469 \\ 0.0250 & 0.3450 & 0.4600 & 0.1450 & 0.0250 \\ 0.0938 & 0.3312 & 0.3375 & 0.1437 & 0.0938 \\ 0 & 0.7500 & 0.2500 & 0 & 0 \end{bmatrix} & P_{ij3} &= \begin{bmatrix} 0.2109 & 0.3922 & 0.2000 & 0.1422 & 0.0547 \\ 0.0980 & 0.4062 & 0.3072 & 0.1452 & 0.0434 \\ 0.0556 & 0.3426 & 0.3985 & 0.1601 & 0.0432 \\ 0.0883 & 0.3464 & 0.3213 & 0.2027 & 0.0414 \\ 0.0938 & 0.3312 & 0.3375 & 0.1437 & 0.0938 \end{bmatrix} \\
 P_{ij4} &= \begin{bmatrix} 0.1545 & 0.3992 & 0.2536 & 0.1437 & 0.0490 \\ 0.0998 & 0.3717 & 0.3222 & 0.1555 & 0.0508 \\ 0.0706 & 0.3561 & 0.3648 & 0.1657 & 0.0428 \\ 0.0874 & 0.3903 & 0.3300 & 0.1490 & 0.0433 \\ 0.0883 & 0.3464 & 0.3213 & 0.2027 & 0.414 \end{bmatrix} & P_{ij5} &= \begin{bmatrix} 0.1271 & 0.3854 & 0.2879 & 0.1497 & 0.0499 \\ 0.0964 & 0.3704 & 0.3251 & 0.1617 & 0.0465 \\ 0.0768 & 0.3661 & 0.3493 & 0.1603 & 0.0445 \\ 0.0925 & 0.3678 & 0.3328 & 0.1581 & 0.0488 \\ 0.0874 & 0.3903 & 0.3300 & 0.1490 & 0.0433 \end{bmatrix}
 \end{aligned}$$

**Step 6.** Predict the value of 22nd time distance and use this value to revise the error between the results and the real value.

The distance value in the 22nd time could be calculated as follow (Table 32.2);

$$T_{22} = \begin{bmatrix} (5.001 + 5.005)/2 \\ (5.006 + 5.010)/2 \\ (5.011 + 5.015)/2 \\ (5.016 + 5.020)/2 \\ (5.021 + 5.025)/2 \end{bmatrix}^T \times \begin{bmatrix} 0.0612 \\ 0.3170 \\ 0.4148 \\ 0.1657 \\ 0.0413 \end{bmatrix} = 5.0120$$

When test the distance value between yellow center and red center in the 22nd time,  $T_{22}$  could be used to compare the result and revises the error between the results and the real value.

**Table 32.2** The corresponding value between times of 17 and 21 to predict the 22 value

Times	Level	Weight	1	2	3	4	5	Source
21	3	0.2314	0	0.2000	0.6000	0.2000	0	$P_{ij1}$
20	4	0.1232	0.0938	0.3312	0.3375	0.1437	0.0938	$P_{ij2}$
19	3	0.3778	0.0556	0.3426	0.3985	0.1601	0.0432	$P_{ij3}$
18	2	0.1968	0.0998	0.3717	0.3222	0.1555	0.0508	$P_{ij4}$
17	1	0.0708	0.1271	0.3854	0.2879	0.1497	0.0499	$P_{ij5}$
Weighted sum $P_i$			0.0612	0.3170	0.4148	0.1657	0.0413	

## 32.4 Experiment Test and Results

### 32.4.1 A Real Example of Judging Deviations

This paper detection method is tested according to the c# program on VS2005 platform. The system adopting this detection method has been applied to a real printing case in Shanghai, China and obtained an effective result. This real application shows that:

This detection method can fast remove interferential image exactly, such as ink spot, noise, nick etc. And the size of the least interferential image that can be detected is 2 pixel;

The detection method ensured the detection reliability, veracity, and real-time. When detecting one frame image whose dimension is 1000\*450 is less than 40 ms, and it can meet max speed achieving approximately 25P/S.

### 32.4.2 Test of Deviation Accuracy and Speed

In order to test the deviation accuracy and speed, the method presented in this paper is compared with the traditional method based on spatial moment algorithm, centroid algorithm and canny operator algorithm in detecting the deviation of Fig. 32.1. For the result fairness, each algorithm tested for 50 times and taking the average experimental result. The result is described as Table 32.3.

From the table it can be seen clearly that traditional method based on centroid algorithm has highest speed, but the positioning accuracy and noise immunity ability are relatively weak; spatial moment algorithm is best in positioning accuracy but lowest speed; canny operator algorithm is weak in both positioning accuracy and speed. While the method presented in this paper has high speed and good positioning accuracy which can match the need of high speed detection of judging printed registration deviation better.

**Table 32.3** Comparison among method based on different algorithm in registration direction

	The method presented in this paper	Centroid algorithm	Spatial moment algorithm	Canny operator algorithm
Calculating time (ms)	39	30	90	62
Accuracy (pixel)	0.32	0.92	0.16	0.78
Anti-noise property	Strong	Weak	Strong	Normal

## 32.5 Conclusion

In order to meet the actual requirements of judging four-color printed registration deviations quickly and accurately, this paper describes a new detection method based on machine vision. This method involving three processes: remove the interferential image of the registration marks firstly by region character separation (RCS) algorithm and uses improved transect interpolating function to detect the edges of mark. Then use weighted markov chain to calibrate the detection. The experiment indicates that the speed and accuracy with this method have greatly improved, and even with noise interference, this method can detect deviation quickly and accurately, superior to the traditional detection method.

**Acknowledgments** This research is supported by the project ansorpts the imported technology and innovation plan in ShangHai (Grant No.11XI-32). The Natural Science Foundation of China (Grant No.61074032). Science and Technology Commission of Shanghai Municipality under Grant No. 10JC1405000.

## References

1. Baochang L, Tong W, Zheng B (2010) An analytical method of updating the range derivatives and a simple image registration method for the MSR-based range Doppler algorithm. *IEEE Geosci Remote Sens Lett* 7(4):831–835
2. Shang H-C, Chen Y-P, Yu W-Y et al (2007) Online autodetection method and system of presswork quality. *Int J Adv Manuf Technol* 33:756–765
3. Vijayaraghavan V, Leung H (2005) A novel chaos based high resolution imaging technique and its application to through-the-wall imaging. *IEEE Signal Process Lett* 12(7):528–531
4. Mumtaz M, Bin Mansoor A, Masood H (2010) A new approach to aircraft surface inspection based on directional energies of texture. In: *International Conference on Pattern Recognition (ICPR)*, 2010, pp 4404–4407
5. Xiangqian P et al (2010) An intelligent online presswork defect detection method and system. In: *Second international conference on Information Technology and Computer Science (ITCS)*, 2010, pp 158–161
6. Yang O et al (2007) A real-time vision system for defect detection in printed matter and its key technologies. In: *Industrial electronics and applications*, 2007, pp 2157–2161
7. Cox IJ, Kruskal JB, Wallach DA (1990) Predicting and estimating the accuracy of a subpixel registration algorithm. *IEEE Trans Pattern Anal Mach Intell* 12(8):721–734
8. Zhou L-h, Xie D-w (2011) Predictions for monthly tripping quantity of circuit breaker in electric distribution network based on weighted markov chain. In: *International conference on Electric Information and Control Engineering (ICEICE)*, 2011

# Chapter 33

## Community Discovering Based on Central Nodes of Social Networks

Ping Fang, Fenglong Shi, Yang Chen and Wanchun Gao

**Abstract** Based on the new concept of central network and the node similarity definition, a fast algorithm is proposed for discovering community structures in social networks. Firstly, start from the maximum degree node, then the two nodes with the maximum number of shared neighbors are taken as the initial community. Next, a neighboring node is judged to push into the initial community according to the appearance frequency of the community it belongs to. Finally, the above step is repeated until all the nodes are classified to the proper communities. The experimental results on two real-world networks demonstrate that the proposed algorithm is able to discover community structure from a given network efficiently and accurately without specifying the community number. The algorithm has a time complexity of only  $O(n)$ .

**Keywords** Social network · Community discovering · Shared neighbours

### 33.1 Introduction

As a kind of complex network, social network represents the relationships of users in social affairs. A social network comprises nodes and edges. The node represents an entity, which usually is an individual or organization. The edge represents the

---

P. Fang (✉) · F. Shi · Y. Chen · W. Gao

Naval Aeronautical and Astronautical University Qingdao Branch, Qingdao 266041, China

e-mail: ping@hust.edu.cn

F. Shi

e-mail: fl-shi@sohu.com

Y. Chen

e-mail: seemeinwinter@sohu.com

W. Gao

e-mail: gaowanchun@126.com

relationship or interaction between two nodes, such as friend, consanguinity. The References [1, 2] point out the degrees of most of real social network show an exponential distribution, which are called scale-free networks. Many researches indicate that a real social network not only is microcosmic and scale-free, but also exhibits a certain community structure. In social networks, a social community can be seen as a sub-graph such that the edges within sub-graph are denser than the edges between sub-graphs. Structure determines function is the basic viewpoint in system science, so it is valuable to discover those communities which maybe represent some real social groups formed by common interests or background in a social network. Community detection is helpful to provide us a much better understanding about the structural topology of each community and analyse its organization principles.

### 33.2 Related Works

In this section, we review a serial of fast algorithms for discovering community structure, such as CNM, Louvain algorithm and dynamical algorithms.

The FN algorithm proposed by Newman [3] is an amalgamation method similar to agglomerative methods using the modularity measure  $Q$ . The two communities are grouped together that give rise to the maximum increase or smallest decrease in  $Q$  at each step. Each step compares at most  $m$  pairs of groups and requires at most  $O(n)$  time to update the  $Q$  value. The algorithm continues until all the  $n$  nodes are in one group and hence the worst case running time of the algorithm is  $O(n(m+n))$ . The CNM algorithm presented by Clauset et al. [4] is essentially a fast implementation of the FN algorithm. Starting from a set of isolated nodes, the edges of the original graph are iteratively added such to produce the largest possible increase of the modularity of Newman and Girvan at each step. The fast version of Clauset et al. which uses more efficient data structures, has a complexity of  $O(n\log^2 n)$  on sparse graphs.

The Louvain algorithm presented by Blondel et al. [5] is a multistep method based on a local optimization of Newman-Girvan modularity in the neighborhood of each node. After a partition is identified in this way, communities are replaced by super-nodes, yielding a smaller weighted network. The procedure is repeated until modularity (always computed with respect to the original graph) does not increase any further. This method offers a fair compromise between the accuracy of the estimate of the modularity maximum and time complexity. The accuracy is better than that delivered by greedy techniques like the CNM algorithm. The complexity is essentially linear in the number of edges of the graph. Label propagation algorithms [6] uses labels to discover communities and takes an almost linear time. COPRA [7] updates the belonging coefficients of each node by averaging the coefficients from all its neighbors in a synchronous fashion. The time complexity is  $O(vm\log(vm/n))$  per iteration. The parameter  $v$  controls the

maximum number of communities which a node can associate with,  $m$  and  $n$  are the number of edges and number of nodes respectively.

Many community detection methods have been proposed in the literature, most of them (1) based on the concept of modularity  $Q$ , but it is also important to note that modularity suffers from resolution limit as shown by Fortunato and Barthelemy, (2) require parameters (which are difficult to quantify) and (3) time complexity is unacceptable when analyzing large scale networks (such as Infomap algorithm [8]).

To overcome these limitations, in this paper, a quick algorithm based on the new concept of central nodes and the node similarity definition is proposed for discovering high quality community structures in social networks. The algorithm thinks the two nodes with the maximum number of shared neighbors more possibly belong to the same community. So these two nodes are taken as the initial community. According to the appearance frequency of the community a neighboring node belongs to, it is judged to push into the initial community one by one to form a new community. Repeat it until all the nodes are in a stable state.

### 33.3 Definitions

In this paper, we focus on simple, undirected and unweighted graphs. Let  $G = (V, E)$  be a graph representing a social network, where  $V$  is the set of  $n$  nodes and  $E$  is the set of  $m$  connections. The node  $v_i \in V$ ,  $C$  is a local community in  $G$ .  $|C|$  is the number of elements in set  $C$ . The following are the concepts and definitions about the proposed algorithm:

#### Definition 1 Central Network

Central network is composed of the nodes whose degrees are greater than the mean node degree in  $G$ . The nodes of central network are defined as central nodes.

#### Definition 2 Neighbor Nodes Set

The neighbor nodes set of node  $i$  is defined:  $N(i) = \{j \mid \text{node } i \text{ links node } j \text{ directly}\}$ . The neighbor nodes set of the local community  $C$  is defined:

$$N(C) = \bigcup_{i=1}^{|C|} N(v_i) - \bigcup_{i=1}^{|C|} v_i, \quad \text{where } v_i \in C$$

#### Definition 3 Node Similarity

The similarity between two nodes is the probability of these two nodes belonging to the same community. Based on the principle of triadic closure, the two nodes quite probably become friends if they share the same friends in the social network. So the number of shared friends is taken to measure the similarity between nodes. On the base of above Definition 2, for  $\forall i, j \in V$ , the node similarity of  $i$  and  $j$  is defined as following:

$$W(i, j) = |N(i) \cap N(j)|,$$

where  $W(i, j)$  is the value after calculation,  $N(i)$  is the set of node  $i$ 's adjacent nodes.

### 33.4 The Algorithm

The algorithm includes the following steps:

**Input:** an undirected and unweighted graph  $G = \langle V, E \rangle$ , let  $i = 1$  and  $j = 1$ .

**Output:** a network community structure.

**Step 1:** create the central network  $G' = \langle V', E' \rangle$  depending on its definition,  $V' = \{v \mid \text{the degree of } v \text{ is greater than the mean node degree in network } G\}$ .

**Step 2:** if  $|V'| > 0$ , find  $v_a$  and its neighbor nodes set  $N(v_a)$  with the maximum degree in  $V'$ .

**Step 3:** find the object node by the Definition 3 in  $N(v_a)$ . Let the initial community  $D_i = v_b + v_a$ ,  $V' = V' - D_i$ ,  $i = i + 1$ , if there is an unclassified node  $v_b$  which has the largest similarity compared to the node  $v_a$ . Whereas let the initial community  $S_j = v_a$ ,  $V' = V' - S_j$ ,  $j = j + 1$ ; return Step 2.

**Step 4:** let  $R = \text{RandomlySort}(V - V')$ , then create the node set  $X = \text{Reverse}(\bigcup_{i=1}^p D_i) \cup \bigcup_{j=1}^q S_j \cup R$ ,  $|V'| = q + 2 * p$ ,  $q \geq 0$ ,  $p > 0$ . Let the initial community  $R_{|V'|+i} = r_i, r_i \in R$ .

**Step 5:** for each node in  $X$ , calculate the appearance frequency  $T$  of the community whose nodes neighboring to it in  $G$ . The community with the highest frequency  $D_{\max(T)}$  is taken as its community. If there are several such communities, the node is push into such community in which the node linking it has the largest degree. Repeat Step 5 until all nodes are steady.

**Step 6:** output the result.

In Step 1, the time complexity to create central network is  $O(n)$ , where  $n$  is the number of nodes in network  $G$ . The time complexity to find the node with the maximum degree in central network is  $O(n')$ , where  $n'$  is the number of nodes in central network and  $n' \leq n$ . In step 3, the time complexity to find the two node with the largest similarity is also  $O(n')$ . The time complexity of Step 5 is  $O(n)$ . So the total time complexity to discover the community structure in the whole network is  $O(n + n' + n' + n) = O(n)$ .



### 33.5 Applications and Result Analysis

In order to evaluate the performance of the proposed algorithm, three typical real-network models are taken as the test networks, whose community structures are already known.

#### 33.5.1 Zachary's Karate Club

In the 1970s, Zachary [9] had observed the social interactions among the members of a karate club at an American university for 2 years, based on which he constructed a social network. The club network consists of 34 nodes and 78 edges, and the network diameter is 5 as shown in Fig. 33.1. As reported by Zachary, the club

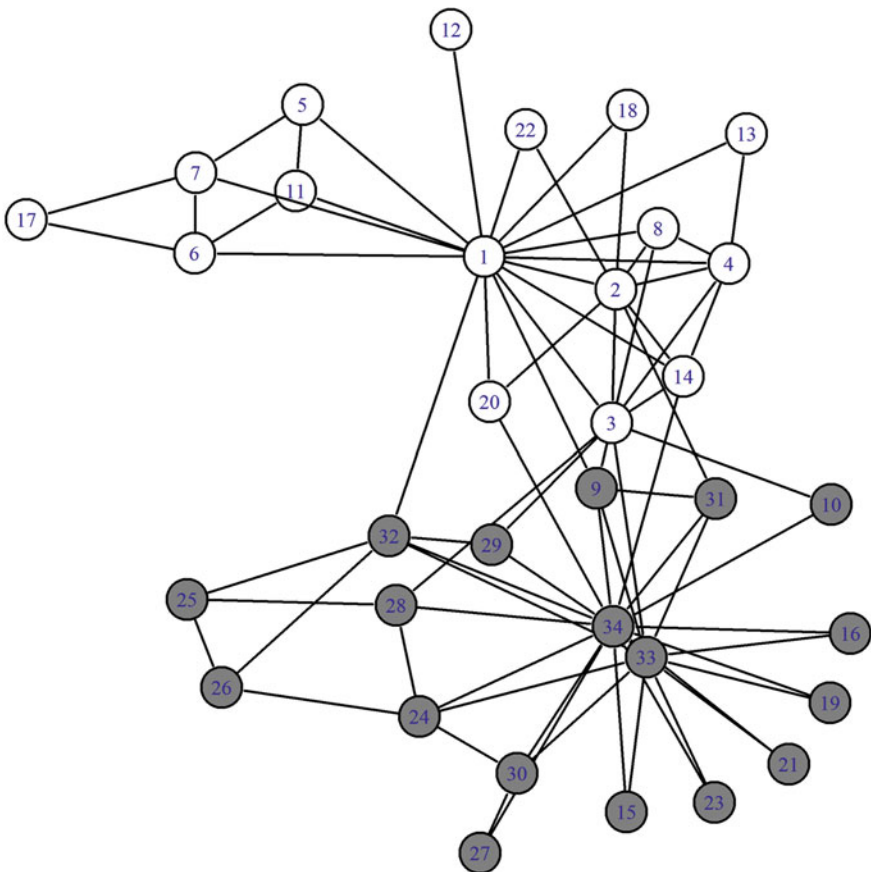
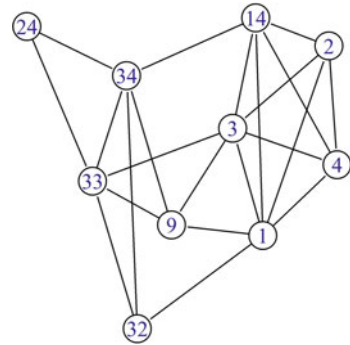


Fig. 33.1 Zachary's karate club network

**Fig. 33.2** The central network of Zachary network



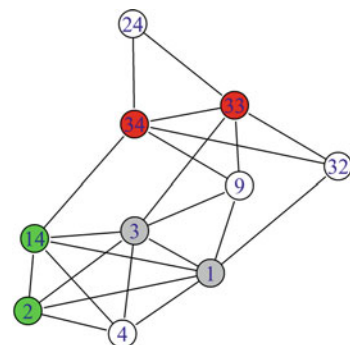
**Table 33.1** The similarity of node 1's neighborhood

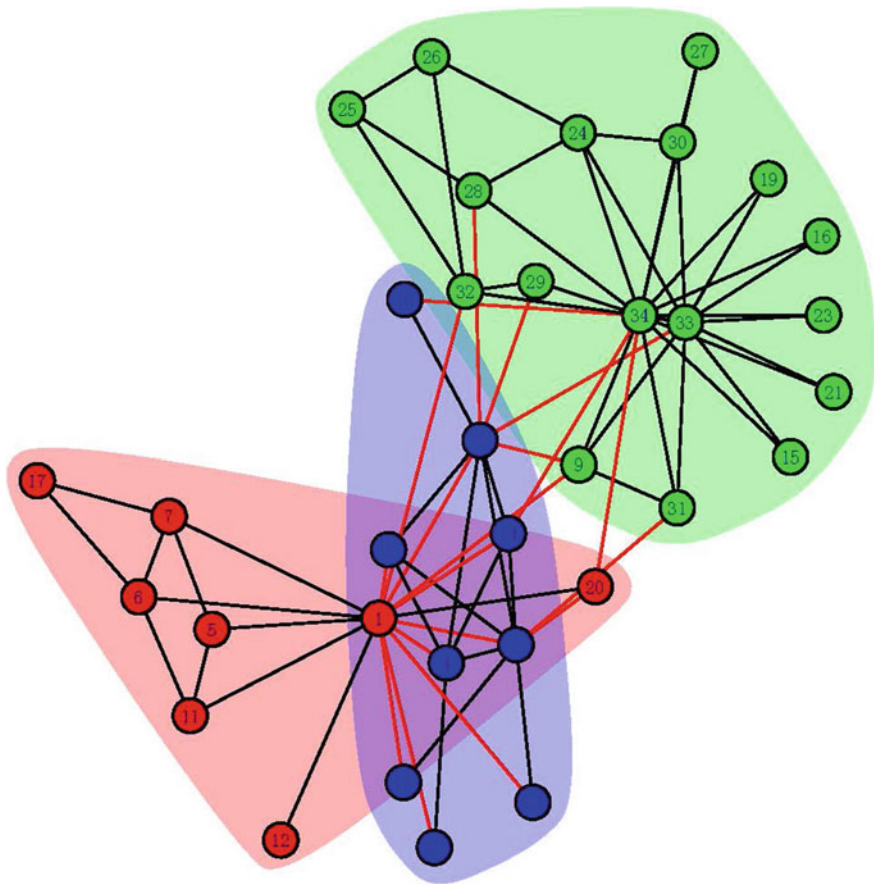
Neighbor	Node similarity
4	3
2	3
14	3
3	4
9	1

eventually splits into two communities due to club expense: Community A denoted by gray circles was led by the administrator (node 33), and Community B denoted by white circles by the instructor (node 1). The Zachary's karate club network has become one of the classic social networks for testing the algorithms of discovering community structures in complex networks.

As shown in Fig. 33.2, the proposed algorithm creates the central network of Zachary network (Zachary central network for short). This central network consists of 10 nodes and 22 edges, in which node 1 and node 3 has the maximum degree. Take node 1 as initial node to find the set of its neighbors and compute the similarity between them as shown in Table 33.1. According to Definition 2, select node 3 with the maximum similarity with node 1 to form the initial community  $D_1 = \{3, 1\}$ . Then repeat the above processing in the remaining nodes until all the

**Fig. 33.3** Community structures identified by the algorithm on Zachary's central network. The communities can be identified by their shades of colors



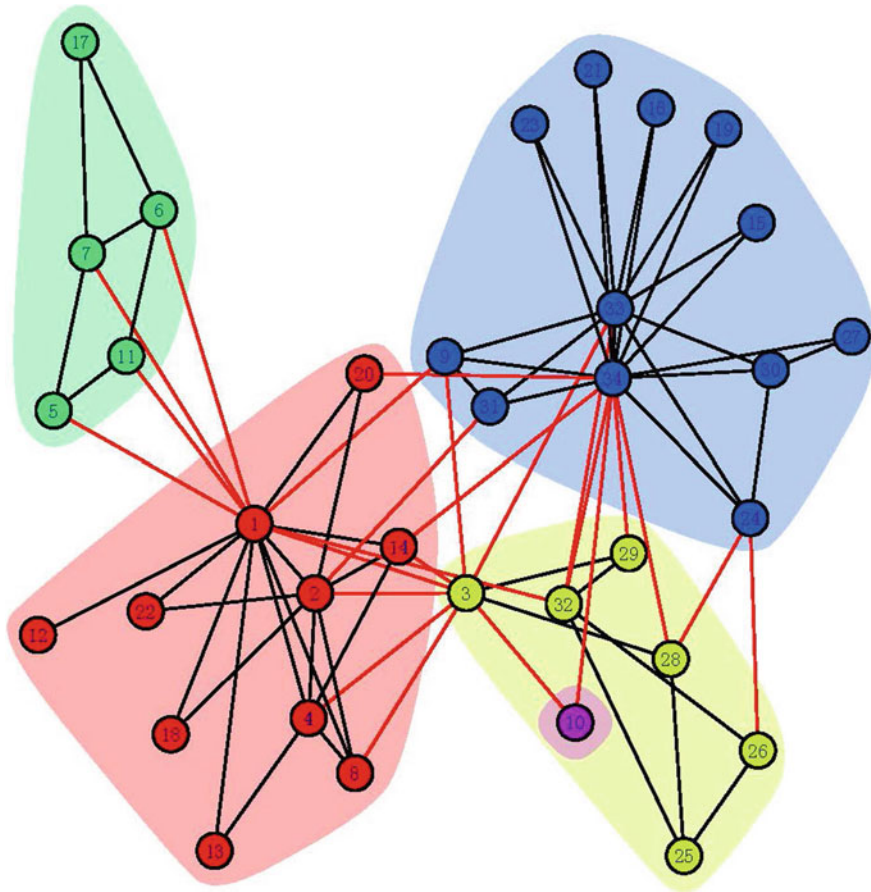


**Fig. 33.4** Communities obtained by CNM

nodes are been classified to its corresponding community. So a serial of initial communities are obtained as shown in Fig. 33.3:  $D_2 = \{2, 14\}$ ,  $D_3 = \{34, 33\}$ ,

$S_1 = \{4\}$ ,  $S_2 = \{9\}$ ,  $S_3 = \{32\}$ ,  $S_4 = \{24\}$ . By Step 4, the node set  $X$  is created. Perform Step 5, the neighbors of each initial community are fast to been classified to the corresponding community. For Zachary network, each node reaches to steady by Step 5 for only one-time performing and the final result is identical with the real community structure observed by Zachary.

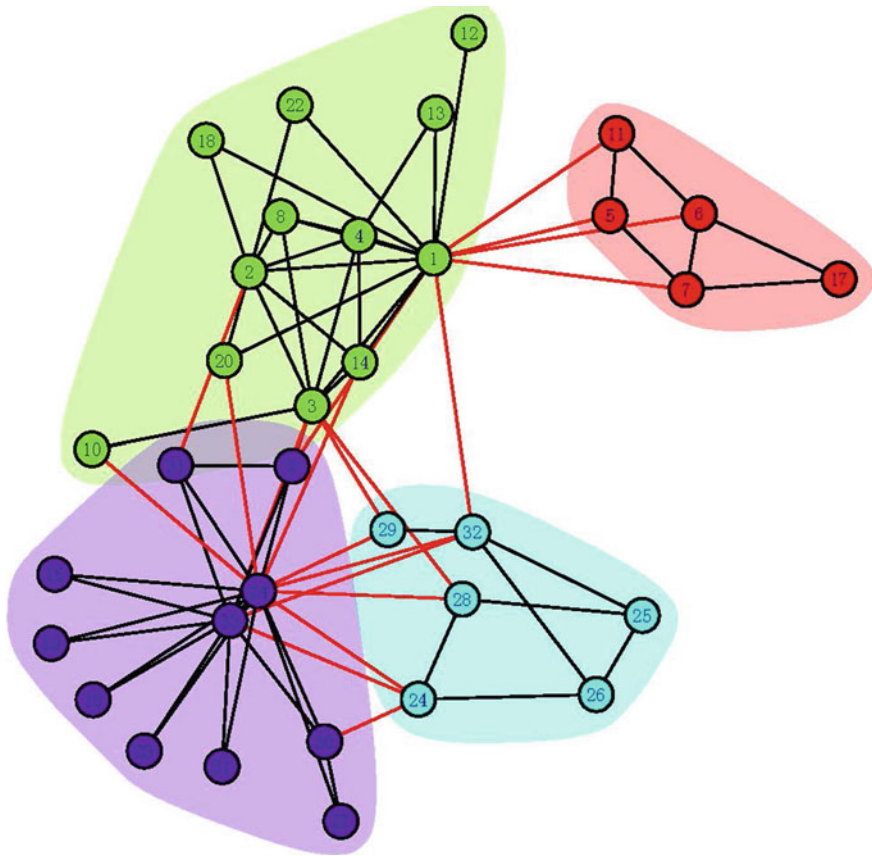
Three results are obtained by the algorithm of LP, in which one of them is identical with the real case. Figure 33.4 shows the community structure obtained by the algorithm of CNM, which comprises three communities. Figure 33.5 shows the community structure obtained by the GN algorithm [10], which comprises five communities. Compared with the above algorithms, the algorithm in this paper is able to discover the community structure of Zachary network effectively without known the number of communities (Figs. 33.6 and 33.7).



**Fig. 33.5** Communities obtained by GN

### 33.5.2 *Dolphin Network*

From 1994 to 2001, Lusseau etc. [11] had observed the social relationship of bottlenose dolphins living in Doubtful Sound of New Zealand. During his research studies, he found these dolphins were separated into two groups for the disappearance of a dolphin (death or leaving) as shown in Fig. 33.8. The value of the modularity  $Q$  for this partition is 0.3735. The node in the network represents a dolphin and an edge connected with two dolphins shows their frequent interactions. A larger group comprises 42 dolphins denoted by the node with white background, whereas a smaller group consists of only 20 dolphins denoted by the node with gray background. This dolphin network contains totally 62 nodes and 159 edges, and the network diameter is 8. The dolphin network is one of the common social networks for analysis.



**Fig. 33.6** Communities obtained by Louvain

For dolphin network, each node reaches to steady by performing Step 5 for two times. Fig. 33.9 shows the final results with three communities, where the nodes with green background represent the smaller group. Compared with the real community structure observed by Lusseau, the node SN89 is classified to the smaller group in error, while others are classified correctly. For the node SN89, its neighbors are Web (its degree is 9) and SN100 (its degree is 7). It is reasonable to classify SN89 to the smaller group based on Step 5. At the same time, the proposed algorithm classifies the larger group more detailedly and gets two communities respectively denoted by the nodes with red and yellow background. In Comparing our results to the results of algorithm GN, CNM, Louvain and Infomap shown in Table 33.2.

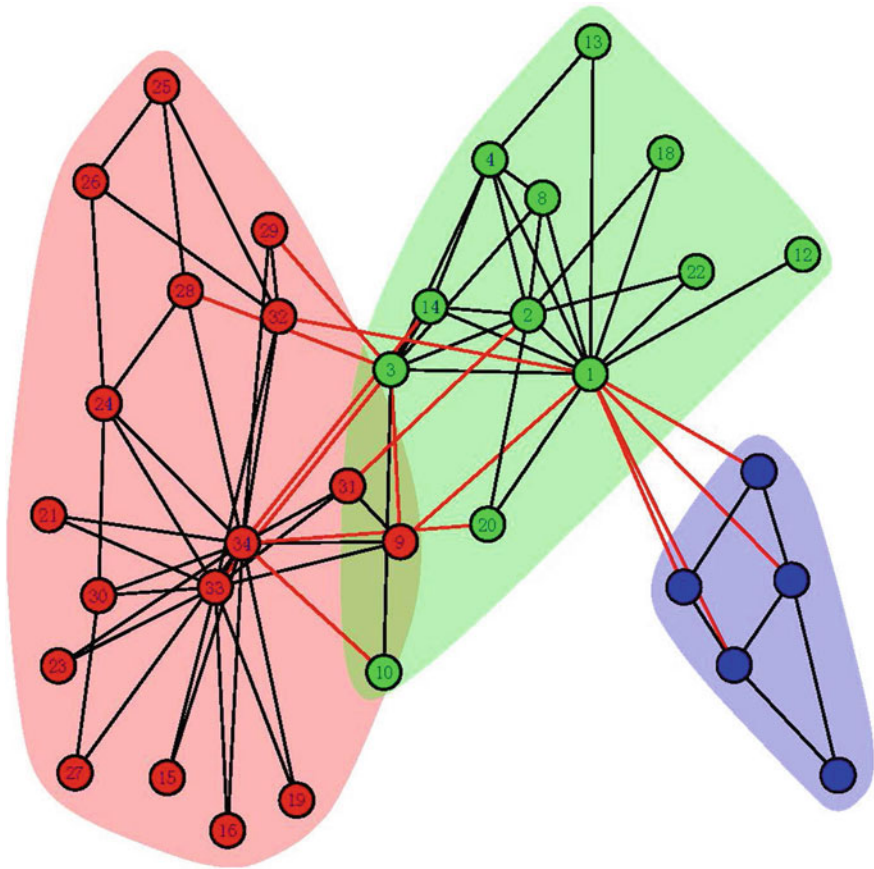


Fig. 33.7 Communities obtained by Infomap

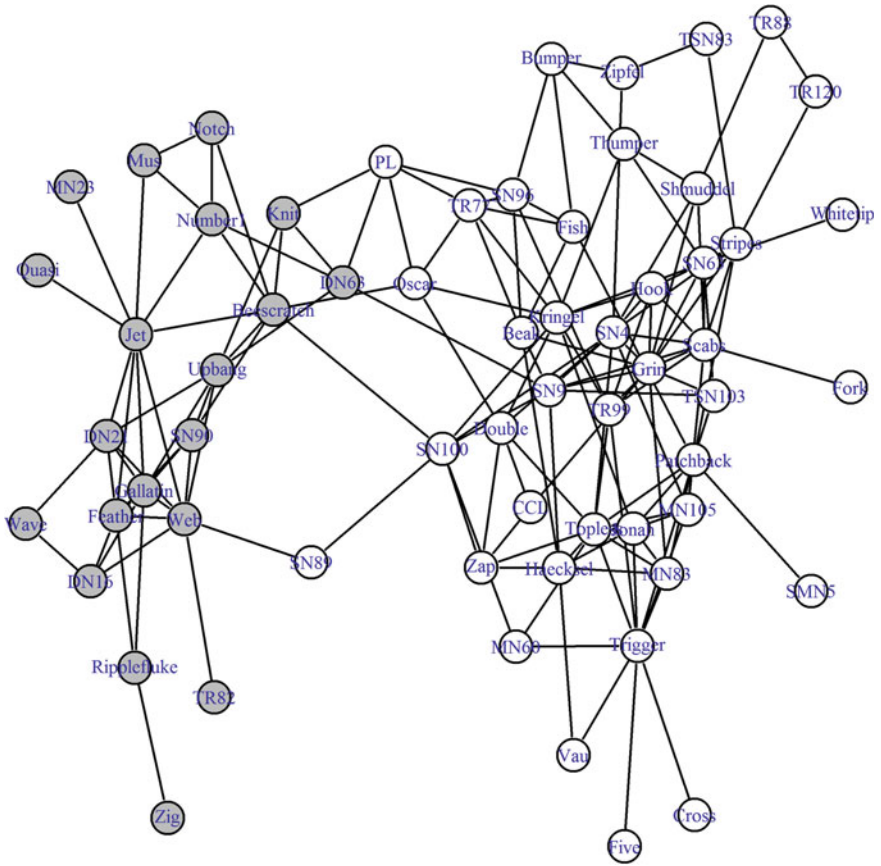
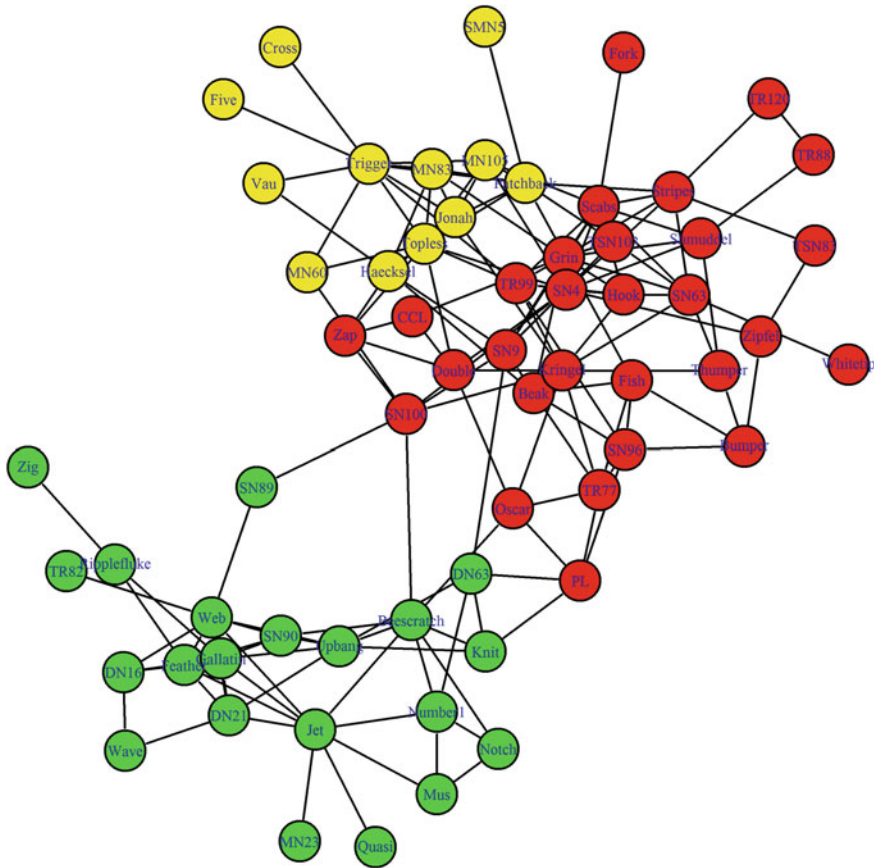


Fig. 33.8 Two groups of the dolphin network



**Fig. 33.9** The final results with three communities

**Table 33.2** Experiments results and comparison

Algorithms	$Q$ value	Cluster sizes
GN	0.5194	1(7), 2(21), 3(20), 4(12), 5(2)
CNM	0.4955	1(22), 2(23), 3(15), 4(2)
Louvain	0.5185	1(7), 2(8), 3(18), 4(14), 5(15)
Infomap	0.5170	1(8), 2(12), 3(11), 4(9), 5(7), 6(5)
Our Algorithm	0.4806	1(21), 2(29), 3(12)

### 33.6 Conclusions

Based on the new concept of central network and the node similarity definition, a fast algorithm is proposed for discovering communities in social networks in this paper. This algorithm is appropriate to those networks with triangle relations. The



result of the proposed algorithm is identical to the real structure of Zachary's karate club network. The result about the dolphin network by the algorithm is rather close to the real structure observed by Lusseau and obtains more details. This algorithm is able to discover the community structure in social network efficiently. The algorithm has a time complexity of only  $O(n)$  and it is easy to implement.

## References

1. Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512
2. Barabási A, Bonabeau E (2003) Scale-free networks. *Sci Am* 288(5):60–69
3. Newman MEJ (2004) Fast algorithm for detecting community structure in networks. *Phys Rev E* 69(6):066133
4. Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. *Phys Rev E —Stat Nonlinear Soft Matter Phys.* 70(62):066111/1–066111/6
5. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *Stat Mech* 10:P10008
6. Raghavan UN, Albert R, Kumara S (2007) Near linear time algorithm to detect community structures in large-scale networks. *Phys Rev E* 76(3):036106
7. Gregory S (2010) Finding overlapping communities in networks by label propagation. *New J Phys* 12:103018
8. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci* 105(4):1118
9. Zachary WW (1977) An information flow model for conflict and fission in small groups. *J Anthropol Res* 33(4):452–473
10. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci U S A* 99(6):7821–7826
11. Lusseau D (2003) The emergent properties of a dolphin social network. *Proc Royal Soc B Biol Sci* 270(Suppl2):186–188

# Chapter 34

## An Improved Force-Directed Algorithm Based on Emergence for Visualizing Complex Network

Hongbo Li, Wenjing Geng, Yu Wu and Xian Wang

**Abstract** Visualization of complex network is one of the most important and difficult issues in complexity science. Most current visualization algorithms are based on drawing aesthetically and it's difficult for them to find the structure information of complex network. To solve this problem, the Fruchterman-Reingold (FR) algorithm based on the force-directed layout is studied in this paper, which is most suitable for the visualization of complex network. From the emergent characteristic of complex network, an improved adaptive FR algorithm is proposed to reduce dependence on parameters in the FR algorithm. In the improved algorithm, the impact of the clustering coefficient on the attraction–repulsion between vertices is considered, and the clustering coefficient is thought to be a determinative indicator for emergence. Then, with the attraction–repulsion the topological characteristic of complex network is visualized. Experiments show that the improved algorithm makes the observation of the structure of complex network much easier. In addition, the improved algorithm displays superior stability and adaptability during experiments.

**Keywords** Emergence · Force-directed algorithm · Complex network · Clustering coefficient · Visualization technology

### 34.1 Introduction

The visualization of complex network is a hot research area in recent years. For instance, the visualization of the routers' topological structure is studied to effectively manage network. The structure and relation of social network are also

---

H. Li · W. Geng · Y. Wu (✉) · X. Wang  
Institute of Web Intelligence, Chongqing University of Posts and Telecommunications,  
No. 2 Chongwen Road, Nanan, Chongqing, People's Republic of China  
e-mail: wuyu@cqupt.edu.cn

visualized to reveal the interpersonal relationship. Indeed, the visualization technology of complex network has played an important role in various fields. For example, in biology, visualization technology has been applied to analyze the interactive network of proteins and gene network [1]. Although visualization technology provides a vivid method for researches on complex network, general visualization algorithms fail to visualize the structure of complex network due to a large number of vertices and complex relationships between them. Therefore algorithms suited to visualize complex network need to be explored and improved according to features of complex network.

There are two thoughts for the visualization of complex network. One is to optimize the layout algorithms and the other is to increase spatial dimensions. Given that layout algorithms are the basis of visualization technology, the first solution is considered in this paper. One of the most famous layout algorithms is the force-directed algorithm (FDA) proposed by Eades [2], also known as the spring-embedded algorithm. FDA is easy to be understood and realized and its graphical layout is beautiful. However, the time complexity of FDA is  $O(N^3)$ , which is very high. Hence, much work has been done to improve FDA's performance [3–5]. And others focus on the aesthetic results by introducing mechanical models. Based on the improved spring-embedded model, Fruchterman and Reingold proposed the FR algorithm [6], where vertices are abstracted as atoms and the graphic theoretic distance between vertices is related to the geometric distance between them in the drawing. Since FDAs mainly imitate the physical system and rarely consider inherent characteristics of the network structure, they need to be improved and optimized when applied to complex network. On the other hand, with the intense development of complex network, increasing researches center on visualization algorithms for complex network. Chan et al. presented out-degree layout algorithm especially for networks with the power law [7]. From the above summarization, it can be well seen that current visualization algorithms for complex network are few and have some limitations.

In addition, the emergent characteristic of complex network attracts more and more attention. “Emergence” emphasizes that by using simple interaction rules, agents at micro level in complex systems organize into an orderly pattern or motion at macro level [8]. In addition to explore emergent phenomena in complex network [9–11], quantitative studies about emergence has been carried out for further prediction and control of emergence. Wu et al. proposed several determinative indicators for emergence [12] and analyzed the universality of those indicators [13]. Those researches provide fresh ideas for the visualization technology of complex network, for both are based on the interaction between agents or vertices.

Therefore, in order to easily observe the structure of complex network, in this paper we choose the FR algorithm as our research basis, and improve it from the perspective of emergence. One reason for choosing the FR algorithm is that with attraction and repulsion between vertices, the system reaches a dynamic balance, which is similar to the mechanism of emergence. Moreover, the clustering coefficient, which is thought to be a determinative indicator, is introduced in the computation of attraction and repulsion between vertices. The experiments with classical

data sets demonstrate that the improvements are effective and the structure of complex network is much easier to observe with the improved algorithm.

## 34.2 The Physical Model of the FR Algorithm and Its Idea

The FR algorithm draws an analogy between edges and forces in natural systems and builds a physical model.

### 34.2.1 The Physical Model of the FR Algorithm

Based on particle physics, each vertex in the graph is abstracted as a particle in the model of the FR algorithm. Since the undirected graph is regarded as a physical system, the positions of vertices are calculated according to the motion of particles in the action of the force field. The physical model of the FR algorithm is given by [6]:

$$\begin{cases} f_a(d) = d^2/k \\ f_r(d) = -k^2/d \end{cases} \quad (34.1)$$

where  $f_a$  and  $f_r$  represent the attractive and repulsive force respectively;  $d$  represents the distance between two vertices;  $k$  is the optimal distance between vertices, namely the radius of a vertex's range of motion. And  $k$  is calculated as:

$$k = C \sqrt{\left(\frac{area}{N}\right)} \quad (34.2)$$

where  $C$  is an experimental constant;  $area$  is the area of the canvas and  $N$  is the number of vertices.

### 34.2.2 The Basic Idea of the FR Algorithm

The FR algorithm considers the attraction and repulsion between vertices to imitate the motion of atoms and celestial bodies. Meanwhile vertices' displacements at every moment are calculated. Then importing temperature variables and cooling function, the positions of all vertices are calculated through multiple iterations with the simulated annealing algorithm.

The basic principle of the FR algorithm is that there exists attraction between vertices with an edge and all vertices repulse each other. More in detail, there are three steps in each iteration. First, calculate the repulsive force between vertices, and determine the displacement vector **disp1** according to the repulsive force.

Then, calculate the attractive force, and calculate the final displacement vector according to the attractive force and  $\mathbf{disp}1$ . Finally, using the cool function, limit the vertices' range of motion.

### 34.3 An Improved Force-Directed Algorithm Based on Emergence

Considering the emergent characteristic of complex network, an adaptive algorithm is proposed based on the following principles.

#### 34.3.1 Principles of Improvement

The current FDAs hardly distinguish the networks' structure, and most improved algorithms firstly take the methods of clustering or dividing communities, and then do placement afterwards. Guided by emergent phenomenon, we aim to put forward a force-directed algorithm which can visualize the intensity of vertices in network with the specific interaction rules. That is to say, vertices in network swarm automatically by setting reasonable interaction rules, to display the networks' structure. Based on the rules proposed above, the improvement principles in this paper are described below.

Firstly, since this improved algorithm is a kind of visualization placement algorithm, it must as far as possible satisfy the five aesthetic criterions proposed by Sugiyama [14]: (1) less-crossings of edges; (2) close layout of vertices connected to each other; (3) straightness of lines; (4) balanced layout of edges; (5) hierarchical layout of vertices.

Secondly, the aim of the improved algorithm is to discover the emergent phenomenon in complex network. In reality, the degree of a vertex in complex network may determine the status and importance of that vertex. In complex network, emergence is related to characteristics of agents. And in the visualization algorithm, the importance of vertices determines their position as well as a special distribution state. So, only the placement algorithm which highlights the different importance between the vertices can helpfully observe emergent phenomena in complex network. Combined with the idea that FR algorithm computes the attraction and repulsion between vertices, we should add the vertices' importance into the computation of attraction and repulsion to make the acting force of vertices conform to the interactive rules between agents in real complex network. However, there are many factors affecting the vertices' importance, such as weight, degree, betweenness, closeness and so on. As the ultimate aim of the improved algorithm is to visualize the complex network, we adopt a universal method to compute the importance, in which the clustering coefficient is introduced to adjust the acting force between vertices in the process of placement every time.

Finally, the ultimate aim of adjusting placement of vertices is to reach a balanced state which can not only fit aesthetic standards but also observe if emergent

phenomenon exists. So we refer the idea of simulated annealing algorithm to cool the energy in each iteration in order to make sure that the system can reach a stable equilibrium state.

### 34.3.2 The Improved Algorithm

The improved algorithm is proposed in the view of emergence, and the physical model is modified to conform to interaction rules in the real world.

#### 34.3.2.1 The Physical Model of the Improved Algorithm

The clustering coefficient in complex network is introduced to calculate the attraction and repulsion between vertices in FR algorithm. The clustering coefficient is calculated as follows [12]:

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \quad (34.3)$$

where  $k_i$  means that there are  $k_i$  edges connected with the vertex  $i$ ;  $E_i$  means that there actually exist  $E_i$  edges between  $k_i$  vertices that are connected with the vertex  $i$ . When  $k_i$  is equal to 0 or 1,  $C_i$  is 0. To avoid this,  $C_i$  is improved by adding 1 and marked as  $NC_i$ :

$$NC_i = \frac{2E_i}{k_i(k_i - 1)} + 1 \quad (34.4)$$

Because the attraction and repulsion exist in every pair of vertices, the vertex's importance is considered in the improved model. The attraction and repulsion between vertices not only depends on their distance, but also is affected by their importance. In this paper, we use  $IM_{ij}$  to represent the importance of the vertex  $i$  and vertex  $j$ , and it is calculated as the following formula:

$$IM_{ij} = \sqrt{NC_i + NC_j} \quad (34.5)$$

After introducing  $IM_{ij}$ , the attraction and repulsion between vertices will change along with the clustering coefficient, which well conforms to the actual characteristics of complex networks. The bigger  $IM_{ij}$  between two vertices are, the bigger the attraction between them and the smaller the repulsion. So, the attraction between vertices should be proportional to  $IM_{ij}$  while the repulsion between them should be inversely proportional to  $IM_{ij}$ . In order to make the cluster more reasonable, the attraction and repulsion are calculated as follows:

$$\begin{cases} f_a(d) = IM_{ij}d^2/k \\ f_r(d) = -\frac{1}{IM_{ij}^2}k^2/d \end{cases} \quad (34.6)$$

where  $f_a$  represents the attraction while  $f_r$  represents the repulsion.

### 34.3.2.2 The Pseudo-code of the Improved Algorithm

The pseudo-code for the improved algorithm is given in the following.

```

G := (V, E); // positions of vertices and edges are initialized.
k :=  $C\sqrt{(\text{width} * \text{height}) / \text{Vnum}}$ ; // width * height is the area of the canvas, and
// Vnum is the number of vertices.

function NC(vertexi) := begin return  $\frac{2E_i}{k_i(k_i - 1)} + 1$  end;

function CalculateIM(vertexi, vertexj) := begin return Math.sqrt(NC(vertexi) +
NC(vertexj)) end;
function  $f_a(d)$  := begin return CalculateIM(vertexi, vertexj) *  $d * d / k$  end;
function  $f_r(d)$  := begin return CalculateIM(vertexi, vertexj) * CalculateIM
(vertexi, vertexj) *  $k * k / d$  end;
//calculate importance between two vertices.
for vertex i in Vertex {
  for vertex j in Vertex {
    if (i != j) {
      CalculateIM(Vertex[i], Vertex[j]);
    }
  }
}
//calculate repulsion and displacement vector caused by repulsion.
while (!end()) {
  for vertex i in Vertex {
    for vertex j in Vertex {
      if (i != j) {
        difference := Vertex[i].pos - Vertex[j].pos;
        Vertex[i].disp := Vertex[i].disp + (difference / |difference|) *  $f_r(x)$ ;
      }
    }
  }
//calculate attraction and displacement vector caused by attraction.
  for edge e in Edge {
    difference := e.source.pos - e.target.pos;
    e.source.disp := e.source.disp - (difference / |difference|) *  $f_a(x)$ ;
    e.target.disp := e.target.disp + (difference / |difference|) *  $f_a(x)$ ;
  }
//limit the range of movement in case of out-of-bounds.
  for vertex v in Vertex {
    v.pos := v.pos + (v.disp / |v.disp|) * min(v.disp, moverange);
    v.pos.x := min(width/2, max(-width/2, v.pos.x));
    v.pos.y := min(height/2, max(-height/2, v.pos.y));
  }
  // reduce the range of movement of each vertex.
  cool();
}

```

The above pseudo-code reflects the core idea of the improved algorithm, and the ending condition of the algorithm is that the displacement of a large number of vertices fluctuates in small scale, which indicates that the total energy of system has reached balance. The time complexity of the improved algorithm is equal to the FR algorithm, which is  $O(N^2)$ . Due to the computation of vertices' importance, the amount of calculation increases a little. But as the clustering coefficient is introduced, the computation of attraction and repulsion between vertices becomes more reasonable. Vertices can cluster automatically according to the change of attraction and repulsion between vertices, which will weaken the dependence on parameters set manually.

## 34.4 Experiments and Analysis

Using three classic network data sets, experiments are carried out to prove the efficiency of the improved algorithm.

### 34.4.1 Visualization Results and Analysis

The FR algorithm is strongly dependent on parameters and has a weak adaptability. A common drawback existed in the force-directed algorithm is that the vertices always concentrate on the canvas's center which leads to hardly distinguish the networks' structure. By contrast, the improved algorithm is less dependent on parameters and has a good adaptability by adding more real interaction rules between vertices, which can reflect characteristics of network's topology in a better way and the groups of tight vertices are easy to recognize. In order to prove this result, some classic data sets like Karate [15], Dolphins [16] and American College football [17] are chosen, and they are given in Table 34.1.

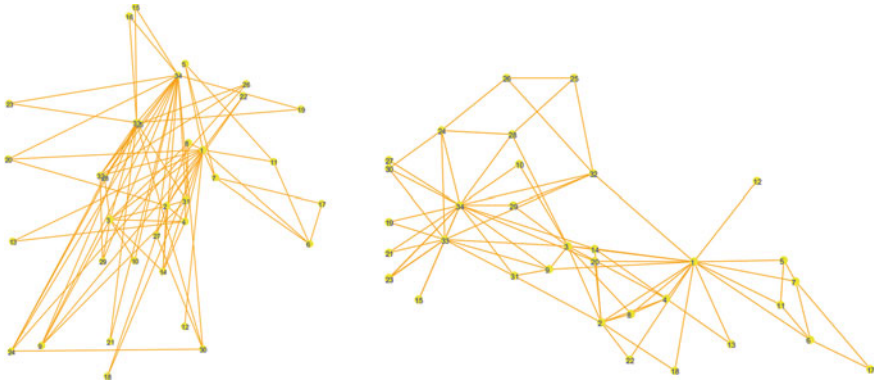
The improved algorithm is developed by JAVA. The area of the canvas is set to 10000\*10000 and  $C$  is set to 1 to make sure vertices uniformly distributed. The visualization results of the above three data sets are shown in Figs. 34.1, 34.2 and 34.3 respectively.

Based on community detection of complex network, Zhu et al. proposed a network topology layout algorithm [18]. The network was divided into several communities by community detecting algorithms and then was drawn by force-

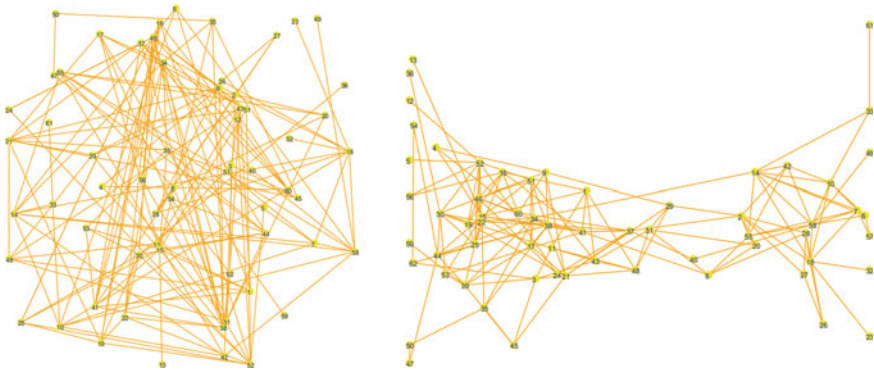
**Table 34.1** Three classic data sets

No	Name of the network	Number of vertices	Number of edges
01	Zachary's karate club	34	78
02	Dolphin social network	62	159
03	American college football	115	616





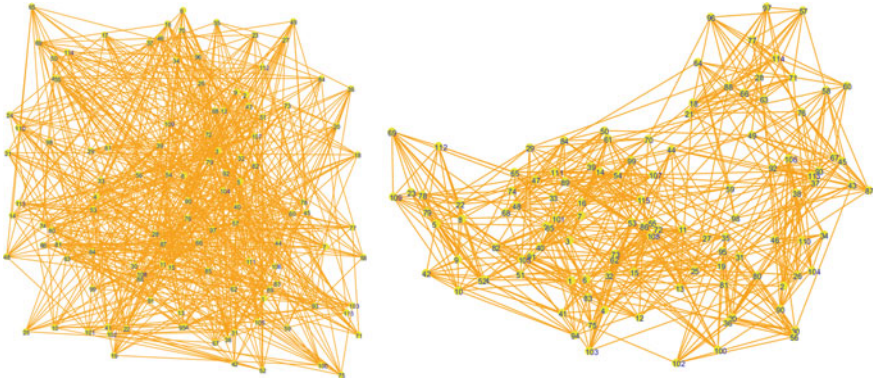
**Fig. 34.1** The Zachary's karate club network. The *left* is the random placement of vertices and edges, while the *right* is the placement with the improved algorithm



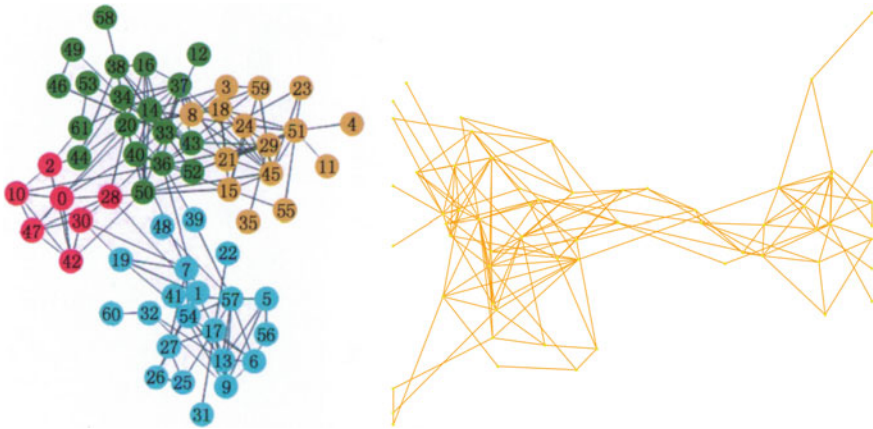
**Fig. 34.2** The Dolphin social network. The *left* is the random placement of vertices and edges, while the *right* is the placement with the improved algorithm

directed algorithm. Since Zhu used the same data sets and aimed at finding structure of network, the visualization results of dolphin social network with two algorithms are compared in Fig. 34.4. It shows that using the improved algorithm, the network topology becomes easier to observe. Although Zhu's algorithm colored the different communities, the full structure is also too tight to recognize the whole topology. The algorithm proposed by Zhu only can be used in the network with less than 3000 vertices. The number of vertices that the improved algorithm deals with is up to about 10000 vertices or more. The more the vertices, the better clustering performance of networks' topology can be observed.

In conclusion, two kinds of comparisons are made in the above. One is the contrast between the initial state and the balanced state with the improved algorithm. And the other is the contrast between the FDA algorithm based on



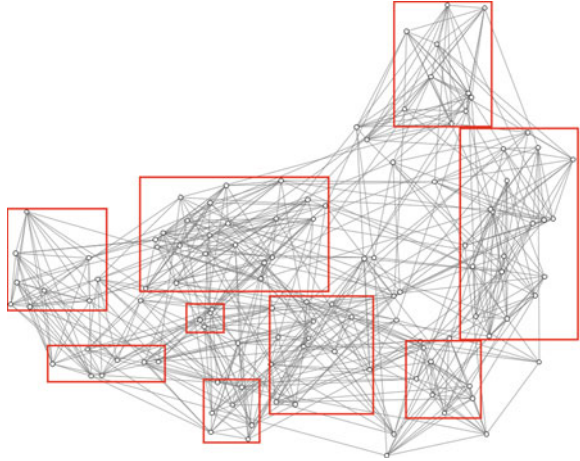
**Fig. 34.3** The American college football network. The *left* is the random placement of vertices and edges, while the *right* is the placement with the improved algorithm



**Fig. 34.4** The comparison between two algorithms. The *left* is drawn by the algorithm proposed by Zhu [18] and the *right* is drawn by the improved method

community division [18] and the improved algorithm based on emergence in this paper. Therefore, it is proved that the improved algorithm contributes to recognize the characteristics of the network's topology in a way of vertices autonomous interaction. As the improved algorithm corresponds to the interactive rules among agents in the real world, it can make a better and more real placement to reflect the structure of networks. Besides, the improved method is only affected by the experimental constant  $C$ , which just affects the moving area of one vertex. That is to say, constant  $C$  only affects the placement's appearance not the structure.

**Fig. 34.5** Skeleton of American college football network



### ***34.4.2 Analysis from the View of Emergence***

The interaction between individuals results in emergence. Based on this, the improved algorithm makes a modification on the previous physical model. And the concept of clustering coefficient is introduced and the interaction rules between vertices are improved. Hence vertices with the high clustering coefficient move closer to each other and then the characteristics of the network topology get to be obvious. It is different from the tradition FDA algorithm which place vertices and edges from the view of aesthetic pleasing.

By setting simple interaction rules, the improved algorithm makes the networks' topology emerge. It can be found that the close vertices may belong to the same club. Drawing the skeleton of American College football network and marking the partly close vertices in red rectangle, it is obviously observed the cluster distribution, as shown in Fig. 34.5. Besides, other social networks, grid networks, traffic networks are also drawn by this algorithm. Experiments show that the vertices display better cluster distributions.

## **34.5 Conclusion**

In this paper, an improved force-directed algorithm based on emergence is proposed. On the one hand, the relationship of vertices and edges can be more clearly and orderly visualized which contributes to observe structure information. On the other hand, combined with the emergent phenomenon, the clustering coefficient is introduced in the visualization algorithm. For this case, we improve the FR algorithm. The results of experiments indicate that the improved algorithm not only amends the randomness of the FR algorithm, but also weakens the dependence on artificially setting parameters for the FR algorithm. What's more, the

improved algorithm can easily observe the topological characteristics in the complex network. Therefore, the improved force directed algorithm based on emergence provides a new method for deeply studying the visualization algorithm for complex network.

**Acknowledgments** Research of this paper is supported by the National Natural Science Foundation of China (No. 60873079, No. 61040044) and Natural Science Foundation of Chongqing of China (cstc2012jjA40027).

## References

1. He S (2010) Study on visualization methods of complex biological networks. Jiangnan University, Wuxi (in Chinese)
2. Eades P (1984) A heuristic for graph drawing. *Congressus Nutnerantiunt* 42:149–160
3. Kamada T, Kawai S (1989) An algorithm for drawing general undirected graphs. *Inf Process Lett* 31:7–15
4. Gajer P, Goodrich MT, Kobourov SG (2004) A multi-dimensional approach to force-directed layouts for large graphs. *Comput Geom* 29(1):3–18
5. Rodgers P, Mutton P (2003) Visualizing weighted edges in graphs. In: Proceedings of the 7th international conference on information visualization. doi:[10.1109/IV.2003.1217988](https://doi.org/10.1109/IV.2003.1217988)
6. Fruchterman TMJ, Reingold EM (1991) Graph drawing by force-directed placement. *Softw Pract Experience* 21(11):1129–1164
7. Chan DS, Chua KS, Leckie C et al (2003) Visualization of power-law network topologies. In: Proceedings of the 11th IEEE international conference on networks. doi:[10.1109/ICON.2003.1266169](https://doi.org/10.1109/ICON.2003.1266169)
8. Wu Y, Tang H, Liu H T (2012) Web swarm intelligence and emergent computation. Science Press, Beijing (in Chinese)
9. Tang H, Wang HT, Huang D (2009) Preliminary research on emergent phenomenon in Internet. *J Chongqing Univ Posts Telecommun (Nat Sci Ed)* 21(5):632–637 (in Chinese)
10. Tang H, Huang D, Wu Y (2010) Overviews on internet emergent behavior research. *Comput Sci* 37(5):34–39 (in Chinese)
11. Wu Y, Yao Y, Wang L (2011) A novel emergence model of public opinion based on small-world network. *J Key Eng Mater* 474–476:2263–2268 (in Chinese)
12. Wu Y, Zhou K, Li YG (2010) The evaluation metrics for swarm emergent behaviors. *Control Theor Appl* 27(8):1086–1092 (in Chinese)
13. Wu Y, Yang JJ, Chen YT (2011) Metrics system of swarm emergence and its efficiency. *J Chongqing Univ Posts Telecommun (Nat Sci Ed)* 23(6):733–740 (in Chinese)
14. Sugiyama K, Tagawa S, Toda M (1981) Methods for visual understanding of hierarchical system structures. *IEEE Trans Syst Man Cybern* 11(2):109–125
15. Zachary WW (1977) An information flow model for conflict and fission in small groups. *J Anthropol Res* 33(4):452–473
16. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci USA* 99(12):7821–7826
17. Lusseau D, Schneider K, Boisseau OJ et al (2003) The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behav Ecol Sociobiol* 54:396–405
18. Zhu ZL, Lin S, Cui K et al (2011) Network topology layout algorithm based on community detection of complex networks. *J Comput Aided Des Comput Graph* 23(11):1808–1815 (in Chinese)

# Chapter 35

## Exploring Efficient Communication in Interactive Dynamic Influence Diagrams

He Wu, Jian Luo and Le Tian

**Abstract** Interactive Dynamic Influence Diagrams (I-DIDs) provide an efficient method for representing multiagent sequential decision problem. By extending I-DIDs with communication, agents are able to exchange their information to learn more about the world. Note that communication is not free, agents should decide whether to communicate or not. This computational process is very time consuming, so it won't work well in a large problem. In this paper, we first study communication based on the framework of I-DIDs, then discuss when agents suppose to communicate considering the cost and the limit resource. Experiments show that our communication algorithm works efficiently in tiger problem. We conclude that communication not only can improve the total rewards, but picking the right time to communicate is also beneficial to agents.

**Keywords** Multiagent systems · Interactive dynamic influence diagrams · Communication decision

### 35.1 Introduction

Partially observable markov decision process (POMDP) has been used successfully for decision making. I-POMDPs [1] extend POMDP to allow agents to use more sophisticated constructs to model and predict behavior of other agents.

---

H. Wu · J. Luo (✉) · L. Tian

Department of Automation, School of Information Science and Technology,  
Xiamen University, Xiamen 361005, China

e-mail: jianluo@xmu.edu.cn

H. Wu

e-mail: wuhe@stu.xmu.edu.cn

L. Tian

e-mail: tianlexm@gmail.com

Interactive dynamic influence diagrams (I-DIDs) [2] are graphical models for sequential decision making in uncertain environment and may be viewed as graphical counterparts of I-POMDPs. I-DIDs generalize dynamic influence diagrams (DIDs) to multiagent settings such as I-POMDPs generalize POMDPs.

Communication is a crucial behavior of agents. Agents in the domain exchange message when they are working on a task, usually the environment is not completely observable for a single agent, then sharing information will allow every agent to learn more. The method of communication we used in this paper is *tell* [3]. The agent at lower level sends its current observation to the agent at higher level.

In this paper, we first introduce the framework of I-DIDs, and extend I-DIDs with communicative act. Then an effective way to communicate is given. Communication is triggered only under some specific cases, that is, when observation is relatively certain. An experiment is made at the end of the paper. It is proved by results to be a very time-saving method, compared to other methods which compute the difference of expected value between communication and non-communication.

## 35.2 Related Works

There already exists some works which discuss on communication in multiagent systems. Xuan et al. introduced a two sub-stage way to complete the communication, and listed a few communication types [3]. Goldman et al. developed a framework which is decentralized semi-Markov decision process with direct communication (Dec-SMDP-Com) [4]. Ghavamzadeh et al. exploited a way to learn to communicate using Hierarchical Reinforcement Learning [5]. Wu et al. proposed a new algorithm called MAOP-COMM in which agents communicate only when inconsistency arises [6]. Carlin et al. presented a theoretical framework to quantify the value of communication and an effective algorithm to manage communication [7]. Roth et al. invented an approach that allows for effective decentralized execution while avoiding unnecessary instances of communication [8]. In another paper, they addressed the question of what to communicate and presented an algorithm that enables multiagent teams to make execution-time decision on how to effectively utilize available communication resources [9]. Nair et al. developed a policy representation that results in savings of both space and time. By communicating every  $K$  steps, even more space and time can be saved [10]. Liu et al. used DAG to maintain and reason the possible joint beliefs of the team at the less memory cost, by which communication decision are made in a decentralized style. Their algorithm is called DAG-DEC-COMM [11]. Also there are some prior work regarding communication based on I-DIDs. Luo et al. proposed a new method by extending I-DIDs with communication abilities, and applied it to solve the path planning problem for AGV in AVS/RS [12]. Zhou et al. introduced a communication decision problem to I-POMDPs, which is called Com-I-POMDPs. They extended the traditional observation function in I-POMDPs and showed the belief update processes [13].

### 35.3 Background

#### 35.3.1 Interactive Influence Diagrams

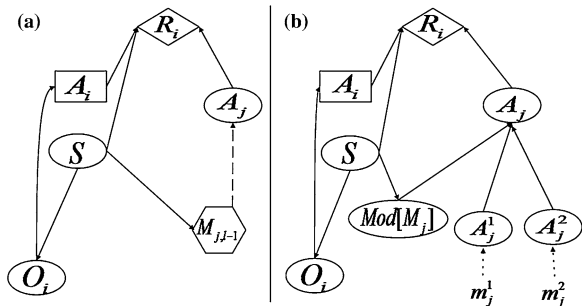
Influence diagrams contain chance, decision, and utility nodes. I-ID is a little different from the normal influence diagram, that is, I-IDs include a new type of node which is named model node. Figure 35.1 shows a general level  $l$  I-IDs with two agents (agent  $i$  and agent  $j$ , for clarity). The model node is denoted by a hexagon. Another difference between I-ID and ID is the “policy link”, which is the dashed link clearly showed in Fig. 35.1a. It presents the distribution over the other agent’s actions given its model.

The model node contains the alternative computational models ascribed by agent  $i$  to the other agent  $j$ . Models in the model node are denoted by  $M_{j,l-1}$ . A model in the model node may be an I-ID or ID or just a simple probability distribution over the actions. The model node and the dashed policy link could be present as shown in Fig. 35.1b, where the model node transforms into chance nodes. Specifically, if  $OPT$  is the set of optimal actions obtained by solving the level  $l-1$  I-ID (or ID), then  $Pr(a_j \in A_j) = \frac{1}{|OPT|}$ , if  $a_j \in OPT$ , 0 otherwise. The conditional probability table (CPT) of the chance node,  $A_j$ , is a multiplexer, that assumes the distribution of each of the action nodes ( $A_j^1, A_j^2$ ), depending on the value of  $Mod[M_j]$  ( $M_{j,l-1}^1, M_{j,l-1}^2$ ). The distribution over the node  $Mod[M_j]$  is the agent  $i$ ’s top-level belief over level  $l-1$  models of  $j$  given the physical state. In the case of more than two agents, we add a model node and a chance node linked by policy link for each other agent.

#### 35.3.2 Interactive Dynamic Influence Diagrams

Interactive Dynamic Influence Diagrams (I-DIDs) extend the formalism of I-IDs to allow sequential decision making over several time steps. We show a general two

**Fig. 35.1** **a** Interactive influence diagrams (I-IDs); **b** A flat ID in order to solve the I-ID



time slice level  $l$  I-DID in Fig. 35.2. Notice that I-DIDs have a special dotted arrow which is called the model update link.

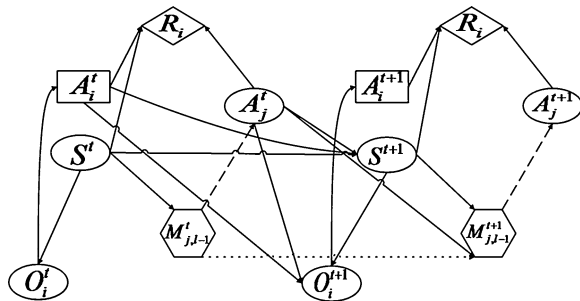
The solution of an I-DID (and I-ID) proceeds in a bottom up manner and is implemented recursively. We start by solving level 0 models which may be traditional DIDs. Their solutions provide the probability distribution which are entered in the level 1 I-DID. The solution method uses the standard look-ahead technique, projection the agent’s action and observation sequences forward from the current belief state and finding the possible belief that  $i$  could have in the next time step. Because agent  $i$  has a belief over  $j$ ’s models as well, the look-ahead includes finding out the possible models that  $j$  could have in the future. Consequently, each of  $j$ ’s level 0 models represented using a standard DID in the first time step must be solved to obtain its optimal set of actions. These actions are combined with the set of possible observations that  $j$  could make in that model, resulting in an updated set of candidate models (including the updated beliefs) that could describe the behavior of  $j$ . Beliefs over these updated sets of candidate models are calculated using the standard inference method through the depending links between the model nodes. Learn more about I-DIDs in the Ref. [2].

### 35.3.3 Communication in I-DIDs

As we mentioned before, communication is a key behavior for agents’ coordination. Agents send their own observations as communication content to each other and these actions don’t change the physical state of the domain. According to [3], before the regular action there is a communication sub-stage. In this sub-stage we have an opportunity to judge if communication should happen or not. We extend I-DIDs to com-I-DIDs with the capability to communicate and name it Com-I-DIDs. We assume the domain has two agents (level 1 agent  $i$  and level 0 agent  $j$ ) for clarity.

As shown in Fig. 35.3, between the regular two time slices in I-DID, there is a middle sub-stage which is represented by dotted lines. The agents’ decision now include communication decision as well. All communications are implemented in a *tell* fashion, that is to say, agent  $j$  will share its own observation with agent  $i$ , and

**Fig. 35.2** An I-DID that is unrolled over two time slice; The dotted arrow between the model nodes is the model update link





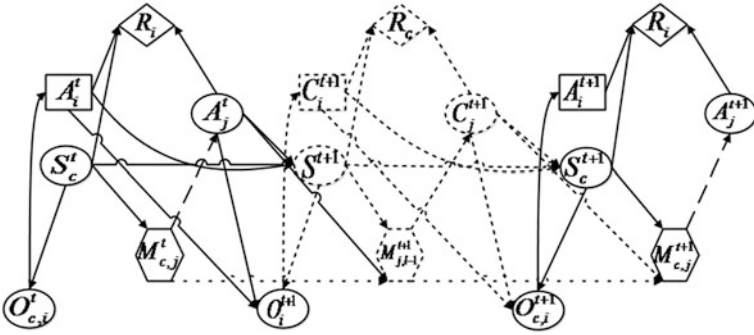


Fig. 35.3 Com-I-DIDs (I-DIDs with sub-stage of communication)

of course, communication is not free. If agent  $j$  choose to communicate, the cost will be  $c(c > 0)$ , or the cost is 0. The communication decision is made in the decision node  $C_j^{t+1}$ . Because communication only effect the state of agents' knowledge and doesn't change the physical states, the CPT of chance node  $S_c^{t+1}$ ,  $\Pr(S_c^{t+1}|S_c^t, c_i^{t+1}, c_j^{t+1}) = 1$ , if  $S_c^t = S_c^{t+1}$ ; 0, otherwise. The value of chance node  $O_{c,i}^{t+1}$ , is agent  $j$ 's observation or null transferred by communication. The CPT of  $o_{c,i}^{t+1}$ ,  $\Pr(o_{c,i}^{t+1} = o_{j,i-1}^{t+1}|s_c^{t+1}, c_i^{t+1}, c_j^{t+1}) = 1$  when agent  $j$  chooses to communicate just after the previous action is finished and before the next action is chosen; Otherwise  $\Pr(o_{c,i}^{t+1} = null|s_c^{t+1}, c_i^{t+1}, c_j^{t+1}) = 1$ . The value in utility node  $R_c$  is the cost of communication. Agents initial communication for the purpose of their own interests. Whether the agent chooses to communicate depends on the terms of the expected utilities of the agent's best actions before and after the communicative act. When the cost for communication doesn't outweigh the expected gain, the agent would perform the communicative act. More information about Com-I-DIDs can be found in the Ref. [12].

### 35.4 Fast Com-I-DIDs

Prevailingly, agents choose whether to communicate depending on the terms of the expected utilities of the agent's best actions with and without the communicative act. When communication brings more expected gain than cost, agent would communicate. This method makes sure the agents gain expected value no less than in the normal case which the agents never communicate.

However, this method deals with a lot of computation so it is very time-consuming. This is because in every time slice it will compare the expected improvement with and without communication (except the first time slice when there is no observation). This means agent must reason about the state and belief of

the other agent twice. Our paper gives a more effective way to communicate. We trigger the communication only when agents' last normal decision brings a relatively more accurate observation. We don't need to compute every time step whether to communicate before the normal phase so that time will be saved. We will explain the method by applying it to the tiger problem [2]. In the typical two-agent tiger problem, we notice that the probability of receiving observation of the world is different. For example, the tiger's location is chosen randomly in the next time step if any of the agents opened any door in the current step. This means agents receive the observation (tiger's growl), "GR" or "GL", 50 % for each. In other case, agent hears the tiger's growls, with the accuracy of 85 %. That is to say, the observation received after the agents perform a "Listen" action is more accuracy than any "Open" actions [14]. Agents sending this observation after "L" seems more meaningful than those that are received after the action "OL" or "OR". Then we use this specific case as a trigger to communicate. After agent  $j$  performs "Listen", it gives its own observation to  $i$  in the same time slice. We denote our algorithm Fast Com-I-DIDs as Com-I-DIDs(F) which is a little different from Com-I-DIDs in Ref. [12]. Another difference between the two is the direction of message passed. In [12], message is passed bidirectionally, but in Com-I-DIDs we use here message is passed from the lower lever agent to higher lever one.

### 35.5 Experiment

We implement our algorithm utilizing Hugin Expert 7.0 (Windows 7, dual processor 1.73 GHz, 2 GB memory) and demonstrate the empirical performance on the well known problem domain: the multiagent tiger problem [2]. We first compare the time consumed in these two methods: Com-I-DIDs and our algorithm denoted as Com-I-DIDs(F) for Horizon = 3, 4, 5. Each data here is the average of 20 runs where the true model of the other agent,  $j$ , is randomly picked according to  $i$ 's belief distribution over  $j$ 's model. As shown in Table 35.1, our algorithm spends less time and may be used in more time slice.

Then we compare the expected value gained in three cases. The first is the normal I-DIDs to which communication never happens. The second is our method and the last one is Com-I-DIDs. Here, communication incurs a cost ( $c = -1$ ). Each data here is the average of 20 runs. As shown in Table 35.2, our method gains more value than the normal I-DIDs, but don't outperform Com-I-DIDs which is no doubt a rational method. Our method intends to balance the lost of expected utility with the efficiency of communicating.

**Table 35.1** Runtime for two methods respectively

Horizon	Com-I-DIDs	Com-I-DIDs(F)
T = 3	20.758	4.835
T = 4	43.314	9.908
T = 5	111.354	26.960

**Table 35.2** Expected rewards gained in three methods respectively

Horizon	I-DID	Com-I-DIDs(F)	Com-I-DIDs
T = 3	9.52	18.48	20.33
T = 4	19.11	22.56	26.72
T = 5	24.52	30.32	37.21

## 35.6 Conclusion

I-DIDs provide a rich framework to model multiagent interacting in decentralized environments and the capability of communication in I-DIDs allows agents to coordinate better by exchanging information. However, a key challenge is how to make a communication decision. We develop a efficient communication policy and describe a way to solve the problem about when to communicate. Agents don't need to communicate every time step blindly, especially when communication incurs a high cost. Agents choose to communicate after they perform some specific action which brings a more convinced message. This message will help agents make better communication choice instead of copious computation. We then implement this in terms of multiagent tiger problem and compare it to other algorithms. Experiment shows the utility improvement that communication brings and runtime is extremely shortened. The key to this method is to find the action which will bring more useful and reliable information. Although our method is not the most optimal one by which agents gain the highest rewards, considered the time and space saved, our method is a tradeoff between rewards gained and limit resource.

**Acknowledgments** This work was supported in part by National Natural Science Foundation of China (Grant No. 60975052).

## References

1. Gmytrasiewicz P, Doshi P (2005) A framework for sequential planning in multiagent settings. *J Artif Intell Res* 24:49–79
2. Doshi P, Zeng Y, Chen Q (2009) Graphical models for interactive POMDPs: representations and solutions. *Auton Agent Multi-Agent Syst* 18(3):376–416
3. Xuan P, Lesser V, Zilberstein S (2001) Communication decisions in multi-agent cooperation: model and experiments. In: *Proceedings of the fifth international conference on autonomous agents*, pp 616–623
4. Goldman CV, Zilberstein S (2008) Communication-based decomposition mechanisms for decentralized mdps. *Artif Intell Res* 32:169–202
5. Ghavamzadeh M, Mahadevan S (2004) Learning to communicate and act using hierarchical reinforcement learning. In: *Proceedings of the third international joint conference on autonomous agents and multi agent systems, AAMAS2004*, vol 3, pp 1114–1121
6. Wu F, Zilberstein S, Chen X (2009) Multi-Agent online planning with communication. In: *Proceedings of the 19th international conference on automated planning and scheduling*, pp 321–328

7. Carlin A, Zilberstein S (2009) Myopic and non-myopic communication under partial observability. In: Proceedings of the 2009 IEEE/WIC/ACM international conference on intelligent agent technology, pp 331–338
8. Roth M, Simmons R, Veloso M (2005) Reasoning about joint beliefs for execution time communication decisions. In: Proceedings of the international joint conference on autonomous agents and multi agent systems, pp 786–793
9. Roth M, Simmons R, Veloso M (2007) What to communicate? Execution-time decision in multiagent POMDPs. In: Proceedings of the eighth international symposium on distributed autonomous robotic systems, pp 177–186
10. Nair R, Roth M, Yohoo M (2004) Communication for improving policy computation in distributed POMDPs. In: Proceedings of the international joint conference on autonomous agents and multi agent systems, AAMAS2004, pp 1098–1105
11. Liu H, Qiao L, Hong B, Piao S (2007) Communication decision in decentralized control of coordinated system. In: Proceedings of the IEEE international conference on mechatronics and automation, Harbin, China, pp 2825–2829
12. Luo J, Yin H, Li B, Wu C (2011) Path planning for automated guided vehicles system via interactive dynamic influence diagrams with communication. In: 9th IEEE international conference on control and automation (ICCA2011), Santiago, Chile, pp 755–759
13. Zhou L, Luo J (2012) A communication model for Interactive POMDPs. In: The 7th international conference on computer science and education (ICCSE 2012), Melbourne, Australia, pp 169–174
14. Emery-Montemerlo R (2005) Game theoretic control for robot teams. Carnegie Mellon University

# Chapter 36

## A No-Reference Remote Sensing Image Quality Assessment Method Using Visual Information Fidelity Index

Yu Shao, Fuchun Sun and Hongbo Li

**Abstract** A novel image quality assessment method for remote sensing image is presented in the paper. Blur and noise are two common distortion factors that affect remote sensing image quality. Those two factors influence each other in both space and frequency domain. So it is difficult to objectively evaluate remote sensing image quality while exist these two kinds of distortion simultaneously. In the proposed method, the input image is first re-blurred by Gaussian blur kernels and also re-noised by white Gaussian noise. Then we measure the amount of mutual information loss before and after image filtering and noising. We take the VIF index as a measure of the information loss. The proposed method does not require reference image and can estimate distorted image with both blur and noise. Experimental results of the proposed method compared with other full-reference methods are presented. It is an accurate and reliable no-reference remote sensing image quality assessment method.

**Keywords** Remote sensing image · Image quality assessment · Human visual system · Visual information fidelity

### 36.1 Introduction

In remote sensing imaging, image quality is determined by various distortion factors. Of these factors, blur and noise are the most commonly used physical characteristics. As is well known, they are described by the modulation transfer function (MTF) and noise power spectrum (NPS), respectively. It is greatly affects

---

Y. Shao (✉) · F. Sun · H. Li  
State Key Laboratory of Intelligence Technology and Systems,  
Department of Computer Science and Technology, Tsinghua University,  
1-511, FIT Building, Tsinghua University, Beijing 100084, China  
e-mail: shaoyu2011@foxmail.com

the subsequent image processing and application. Remote sensing image quality analysis not only can be used for guiding on-orbit remote sensing imaging control, but also to make a preliminary assessment of the quality of image, so it has widespread application.

Remote sensing image quality assessment (IQA) can be divided into two methods: subjective evaluation and objective evaluation. Subjective method requires large amount of people in the completely same condition to mark the image and the mean opinion score is used as the final score of the image, which makes it really time-consuming, cumbersome and expensive to conduct for mass remote sensing image data processing. Objective IQA measures aims to predict perceived image quality by human subjects, which are the ultimate receivers in most image processing applications. Depending on the availability of a pristine reference image, which is presumed to have perfect quality, IQA measures may be classified into full-reference (FR), reduced-reference (RR), and no-reference (NR) methods. In the actual application, remote sensing image usually can't get reference image, so NR IQA has great application advantages.

Since the 1970s, US has developed NIIRS standard [1] and GIQE equation [2], which can give image quality evaluation if remote sensor parameters are known or can be obtained. Remote sensing image is ultimately for human visual perception. A large number of studies show that considering the characteristic of the human visual system (HVS) in IQA is better than those that do not consider the HVS. But It is very difficult to make objective evaluation results match human visual perception. In recent years some HVS based evaluation models [3, 4] were proposed, but these evaluation models are mainly for a particular type of image distortion [5–8]. An imaging system may only be superior in one metric while being inferior in other metrics.

In this paper, based on the analysis of the HVS and in-depth understanding of the influence of noise and blur on remote sensing image quality, we proposed a NR remote sensing IQA method based on visual information fidelity (VIF) index called PVIF. Experimental results show that PVIF can well reflect the visual perception of the image quality effect.

The paper is organized as follows. Human visual characteristics are discussed in Sect. 36.2. Section 36.3 presents the proposed NR IQA method. The experiments are analyzed in Sect. 36.4 and conclusions are drawn in Sect. 36.5.

## 36.2 Visual Characteristics of Remote Sensing Image

When we use eyes to observe a remote sensing image, the incentive from image is the combination of signal stimulus with different frequencies and amplitude. The human eye's response to an excitation signal may also be influenced by other incentives. Contrast masking [9] refers to the reduction in visibility of one signal stimulus caused by the presence of another signal stimulus. Due to the existence of visual contrast masking, some distortions of remote sensing image may be ignored



**Fig. 36.1** From *left to right*: original image, original image blurred with an averaging filter, blurred image re-blurred with the same averaging filter

by human eyes. Those distortions will not affect the overall image quality; but another distortion will be strengthened, that seriously deteriorate image quality.

Image blur effect is caused by the loss of the high frequency content. It can be reproduced with a low-pass filter. We observe that it is difficult to perceive differences between a blurred image and the same re-blurred image. If we blur a sharp picture, image quality will change with a major variation. On the contrary, if we blur an already blurred picture, image quality will still change, but only to a weaker extent. In Fig. 36.1, we present from left to right the original sharp image, the original image blurred with a low-pass filter and the blurred image re-blurred with the same low-pass filter. We observe a high difference in term of loss of details between the first and the second image and a slight difference between the second and the third image. We can explain this phenomenon by the fact that the second blurring effect reduces the difference between pixels that has already been reduced by the first blurring effect. If we add noise to an already noised image, due to the existence of visual contrast masking of HVS, we notice that the high differences significantly decrease after the first noising step and slightly decrease after the second noising step as shown in Fig. 36.2.



**Fig. 36.2** From *left to right*: original image, add white Gaussian noise to original image, noised image re-noised by add the same white Gaussian noise

### 36.3 The Proposed Method

The key idea of our IQA principle is to re-blur and re-noise the input image and to analyze the behavior of the mutual information. As one of the most popular FR IQA method, the Visual Information Fidelity (VIF) [10] takes an information theoretic framework of visual contents based on natural scene statistics models. We first give a brief introduction of the VIF index, and then propose our method.

#### 36.3.1 VIF index

Figure 36.3 provides an overview of the VIF. Let  $C$  pass through HVS, which is modeled as a loss channel, and call the output  $E$ . Let a distorted image  $D$  also be subjected to HVS loss, and we refer to the output of HVS channel as  $F$ . In the view of information theory, mutual information between  $C$  and  $E$  ( $I(C, E)$ ) reveals the amount of information that “loss channel” preserves about the input  $C$ . VIF interprets this mutual information as a way to assess image quality in HVS.

Let  $C$  and  $D$  denote the random fields (RFs) from the reference and distorted images respectively.  $C$  is a product of two stationary RFs that are independent of each other:  $C = SU = \{S_k U_k : k \in I\}$ , Where  $I$  denotes the set of spatial indices for the RFs,  $S$  is a RFs of positive scalars, and  $U$  is a Gaussian scalar RFs with mean zero and variance  $\sigma_U^2$ .

The image distortion model is a signal attenuation and additive Gaussian noise, defined as  $D = GC + V = \{g_k C_k + V_k : k \in I\}$ , where  $G$  is a deterministic scalar attenuation field, and  $V$  is a stationary additive zero-mean Gaussian noise RFs with variance  $\sigma_V^2$ .

The human visual system (HVS) model in VIF quantifies the impact of the image that flows through HVS:  $E = C + N$  and  $F = D + N$ , where  $E$  and  $F$  denote the cognitive outputs of the reference and test images extracted from the brain, respectively;  $N$  represents stationary white Gaussian noise RFs with variance  $\sigma_n^2$ .

VIF utilizes mutual information  $I(C_k, E_k)$  to measure the information that can be extracted from the output of HVS when the reference image is being viewed

$$I(C_k, E_k) = \frac{1}{2} \log_2 \left( \frac{|s_k^2 C_U + \sigma_N^2 I|}{|\sigma_N^2 I|} \right) = \frac{1}{2} \log_2 \left( 1 + \frac{\sigma_{C_k}^2}{\sigma_N^2} \right). \quad (36.1)$$

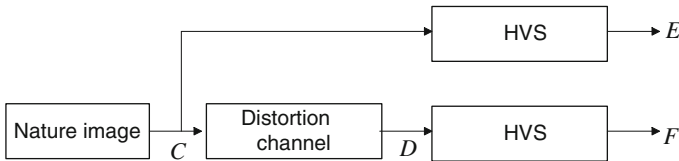


Fig. 36.3 VIF flow diagram



In addition, information  $I(C_k, F_k)$  is measured in the same way when the test image is being viewed

$$I(C_k, F_k) = \frac{1}{2} \log_2 \left( \frac{|g_k^2 s_k^2 C_U + (\sigma_N^2 + \sigma_{V_k}^2) I|}{|(\sigma_N^2 + \sigma_{V_k}^2) I|} \right) = \frac{1}{2} \log_2 \left( 1 + \frac{g_k^2 \sigma_{C_k}^2}{\sigma_N^2 + \sigma_{V_k}^2} \right) \quad (36.2)$$

Also, we have only dealt with one sub-band so far. One could easily incorporate multiple sub-bands by assuming that each sub-band is completely independent of others in terms of the RFs. The VIF index assesses mutual information between  $C$  and  $E$  (and  $C$  and  $F$ ) as follows:

$$\text{VIF}(C, D) = \frac{\sum_{j \in \text{subbands}} \sum_k I(C_k^j, F_k^j)}{\sum_{j \in \text{subbands}} \sum_k I(C_k^j, E_k^j)} = \frac{\sum_j \sum_k \log_2 \left( 1 + \frac{(g_k^j)^2 \sigma_{C_k^j}^2}{\sigma_N^2 + \sigma_{V_k^j}^2} \right)}{\sum_j \sum_k \log_2 \left( 1 + \frac{\sigma_{C_k^j}^2}{\sigma_N^2} \right)} \quad (36.3)$$

### 36.3.2 The Proposed NR IQA Based on VIF Index

In order to measure the quality of remote sensing images, we first obtain a re-blurred image by filter the input image with Gaussian kernel low-pass filter, then add white Gaussian noise to input image and get a re-noised image. As we discussed in the previous section, we measure the amount of information changes before and after image filtering and noising. We take the VIF index value as a measure of this information changes. Bigger VIF values represent smaller image information changes. We obtained the final IQA results by combining information changes at each pixel. Fig. 36.4 shows a flowchart of the proposed NR IQA algorithm, and the whole steps are as follows.

Step 1: A re-blurred image  $I_b$  is produced by applying Gaussian filter to the input image  $I_0$ .

Step 2: A re-noised image  $I_n$  is produced by adding white Gaussian noise to the input image  $I_0$ .

Step 3: Compute  $\text{VIF}(I_0, I_b)$  and  $\text{VIF}(I_0, I_n)$  using Eq. (36.3).

Step 4: The  $\text{VIF}(I_0, I_b)$  and  $\text{VIF}(I_0, I_n)$  are used to construct the final IQA index by compute  $\text{PVIF}(I_0) = (1 - \text{VIF}(I_0, I_b))(1 - \text{VIF}(I_0, I_n))$ .

## 36.4 Experimental Results

In this section, we test the proposed NR IQA method on some remote sensing images, compared with FR-IQA method VIF index and PSNR. We use 4 typical remote sensing images from worldview-2 satellite as test images shown in

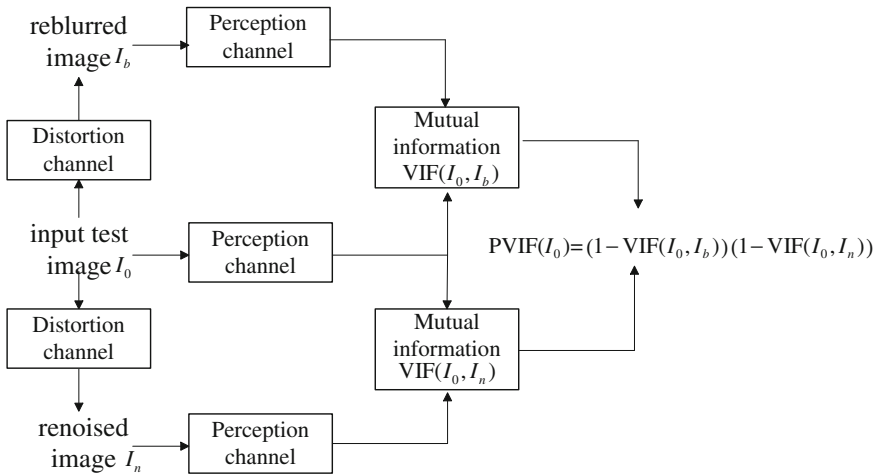


Fig. 36.4 Flowchart of proposed algorithm

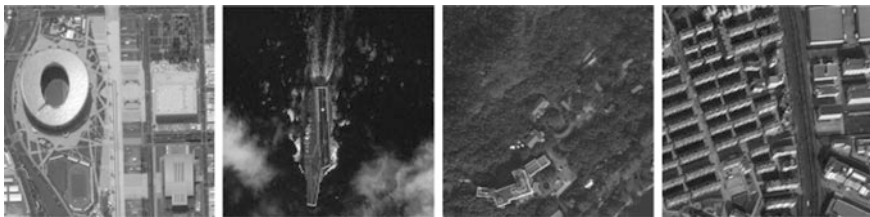


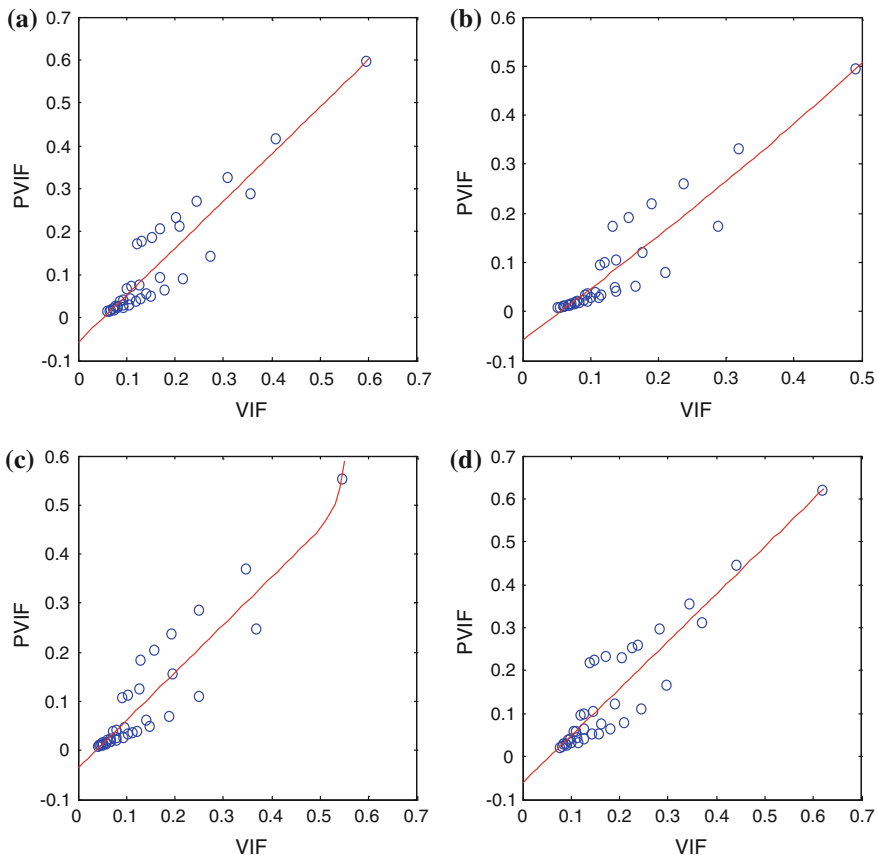
Fig. 36.5 Test images used in the experiments

Fig. 36.5. Then we use these 4 test images to generate a series of distorted images denoted as C1, C2, C3 and C4 respectively by adding different Gaussian blur and Gaussian noise. Spearman rank-order correlation coefficient (SROCC) [11] is used to assess performance of the quality index.

Table 36.1 lists the SROCC results of PVIF and the two IQA algorithms on the C1, C2, C3 and C4. From Table 36.1, we can see that the proposed VIF based IQA metric PVIF performs consistently well across all the test images. The SROCC between PVIF and VIF are greater than 0.92 in all 4 test images, and exceed 0.6 when compared with PSNR. This means when the reference image does not exist, PVIF can replace VIF and PSNR and give a more accurate quality evaluation. Figure 36.6 shows the scatter distributions of VIF versus the predicted scores by

Table 36.1 The SROCC performance of PVIF compare with VIF and PSNR

PVIF	C1	C2	C3	C4
VIF	0.932	0.928	0.925	0.912
PSNR	0.726	0.646	0.611	0.630



**Fig. 36.6** Plots of VIF versus PVIF of the image sets degraded from 4 test images

PVIF. The curves shown in Fig. 36.6 were obtained by a nonlinear fitting [12]. From Fig. 36.6, one can see that the objective scores predicted by PVIF correlate consistently with the VIF index.

### 36.5 Conclusion

We have presented a novel robust, low-cost no-reference remote sensing image quality assessment algorithm. The mutual information between an original image and its re-blurred and re-noised versions has been proposed to estimate image quality. The proposed method has been shown to have robust estimation.

**Acknowledgments** This study was funded by National Basic Research Program of China (973 Program) under Grant 2012CB821206.

## References

1. Jon CL (2003) Image quality equation and niirs. *Encycl Opt Eng* 1:794–811
2. Thurman ST, Fienup JR (2008) Analysis of the general image quality equation. In: *Proceedings of SPIE*, (6978), 69780F
3. Zeng Y, Wang W (2012) Optimal display of remote image based on hvs and its applications. *Spacecraft Recovery Remote Sens* 1(33):46–52 (in chinese)
4. Zhang F, Xie W, Lin L, Qin Q (2011) No-reference remote sensing image quality assessment based on natural scene statistical in wavelet domain. *J Electron and Inf Technol* 11(33):2742–2747 (in chinese)
5. Cohen E, Yitzhaky Y (2010) No-reference assessment of blur and noise impacts on image quality. *Signal Image Video Process* 3(4):289–302
6. Wang Z, Xie Z, He C (2010) A fast quality assessment of image blur based on sharpness. In: *3rd international congress on image and signal processing (CISP)*
7. Xin W, Baofeng T, Chao L, Dongcheng S (2008) Blind image quality assessment for measuring image blur. In: *2008 congress on image and signal processing(CISP 2008)*
8. Li C, Yang X, Chen W, Lu W (2009) Study on the iqa method for polarization image based on degree of noise pollution. In: *International conference on information and automation*
9. Wang Z, Bovik AC (2009) Mean squared error: love it or leave it? A new look at signal fidelity measures. *IEEE Signal Process Mag* 1(26):98–117
10. Sheikh HR, Bovik AC (2006) Image information and visual quality. *IEEE Trans Image Process* 2(15):430–444
11. Zhang L, Zhang L, Mou X, Zhang D (2012) A comprehensive evaluation of full reference image quality assessment algorithms. In: *The international conference on image processing*
12. Sheikh HR, Sabir MF, Bovik AC (2006) A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans Image Process* 11(15):3440–3451

# Chapter 37

## 3D Model Feature Extraction Method Based on the Partial Physical Descriptor

Kuansheng Zou, Haikuan Liu, Zengqiang Chen and Jianhua Zhang

**Abstract** With the rapid development of 3D scanners, graphic accelerated hardware and modeling tools, the application of 3D model databases is growing in both numbers and size. There is a pressing need for effective content-based 3D model retrieval methods. In this paper, a novel 3D model retrieval system called Physical Descriptor (PDD) is proposed. The physical descriptor is defined as the physical features extracted from the 3D surface. Firstly, after pose normalization for 3D database, the 3D model is partitioned into several parts by the planes paralleling the XOY, YOZ and XOZ plane respectively. Each partial part is represented by a physical feature named as PPD, which is a combination the inertia moment, elastic potential energy and the density of the sliced part. Several retrieval performance measures demonstrate that the proposed approach is superior to other methods.

**Keywords** 3D model retrieval · Partial physical descriptor · Continuous principal component analysis · Feature extraction

---

K. Zou (✉) · H. Liu · J. Zhang  
School of Electrical Engineering and Automation, Jiangsu Normal University,  
Xuzhou 221116, China  
e-mail: zoukuansheng@163.com

H. Liu  
e-mail: lhkeed@xznu.edu.cn

J. Zhang  
e-mail: zhangjianhua@jsnu.edu.cn

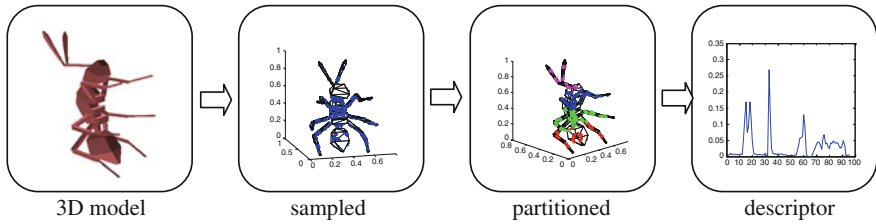
Z. Chen  
College of Information Technical Science, Nankai University, Tianjin 300071, China  
e-mail: chenzzq@nankai.edu.cn

## 37.1 Introduction

Duplicate designs consume a large amount of enterprise resources during product development. Automatic search for similar parts is an effective solution for design reuse [1]. The popularization of 2D image retrieval and 3D data development had created an urgent demand for more effective 3D model retrieval systems. Since searching for similar models in 3D databases can sometimes speed up the design and research processes, thus, the content-based 3D model retrieval methods have become a hot topic in the last few years [2]. Such methods can be used to discover geometric relationships between 3D data, and find the required data from the local databases or from the Internet.

The commonly adopted methods are distribution-based [3–6] and 2D views-based [7, 8]. Distribution-based approaches rely on the idea of accumulating feature information to obtain a global shape description. The D2 descriptor [3], which is a probability distribution histogram of two randomly selected points from the model's surface, is robust against the degeneracy of 3D models. However, it sacrifices the discriminative accuracy. A fractal D2 (FD2) descriptor is proposed to improve the performance of D2 [4]. FD2 is a two dimensional distribution by using D2 and 3D fractal dimensions. Angle Distance (AD) and Absolute Angle Distance (AAD) descriptors are proposed to compare 3D models [5]. AAD measures the distribution of absolute angles between the normal vectors of two associated surfaces, where the randomly selected points are located. Then it is combined with the distance of two selected points. It is a two dimensional descriptor which contains both the distance and the angle information. In this research, an exhaustive study of second order 3D shape features has been carried out. It was found that many combined shape descriptors were proposed based on group integration, such as Beta/Distance (BD) and Alpha/Beta/Distance (ABD), and experiments showed that further improvements of shape distributions can also lead to better results than the well known methods [6].

2D views-based methods consider the 3D shape as a collection of 2D projections taken from different view points of the 3D model, and each projection is then described by standard 2D image descriptors, such as Fourier descriptors or Zernike moments. These methods can obtain a good retrieval performance but have large feature size and high matching cost. Chen et al. [7] proposed the Light Field Descriptor (LFD), which is comprised of Zernike moments and Fourier coefficients computed on a set of projections taken from the vertices of a dodecahedron. Vranic [13] proposed a shape descriptor where features are extracted from depth buffers produced by six projections of the object, one for each side of a cube which encloses the object. In the same work, the Silhouette-based (SIL) descriptor is proposed which uses the silhouettes produced by the three projections taken from the Cartesian planes. Vranic [8] developed a hybrid descriptor called DESIRE, which consists of the Silhouette, Ray and Depth buffer based descriptors, which are combined linearly by fixed weights.



**Fig. 37.1** Feature extraction of the PPD descriptor

In this research, a novel 3D model feature extraction method is proposed by using the partial physical descriptor (PPD). After the preprocessing for the 3D models, each model is partitioned into several parts by the planes paralleling the XOY, YOZ and XOZ planes respectively. Each sliced partial part is represented by a physical feature named as PPD, which is a combination of the inertia moment, elastic potential energy and the density.

## 37.2 Feature Extraction of Partial Physical Descriptor

The concept of the PPD is that if two 3D models correspond to each other, then their sliced parts should also correspond to each other. The flow chart of the PPD extraction is shown in Fig. 37.1.

The 3D model is sampled and partitioned into many regular parts at first, and then the Inertia Moments Descriptor (IMD), the Potential Energy Descriptor (PED) and the Density Descriptor (DD) are computed respectively; finally the PPD is obtained by combining the three descriptors. The detail steps are shown as follows.

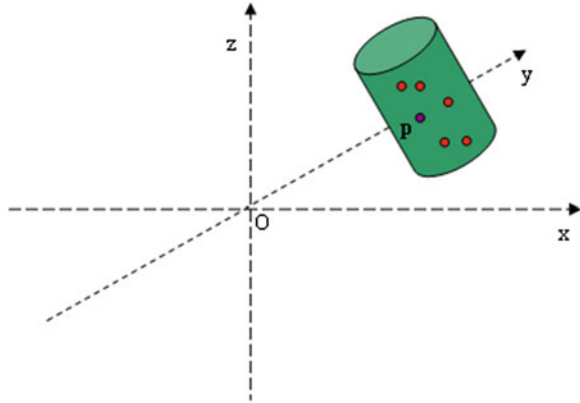
### 37.2.1 Definition of PPD

Assume that  $S$  is a sliced part of a 3D model, it can be seen as an elastic part, which has the elastic potential energy, and its rotation characteristic is described as the inertia moments. Assume that it contains  $k$  random sampled points, the total features combined together called the partial physical descriptor (PPD) (Fig. 37.2).

### 37.2.2 Model Partitioning Based on 3D Pose Estimation

The pose estimation procedure initially involves the translation and scaling of the 3D model. The model is translated so that the center of mass coincides with the

**Fig. 37.2** A sliced part in the 3D space, which contains  $k$  randomly sampled points

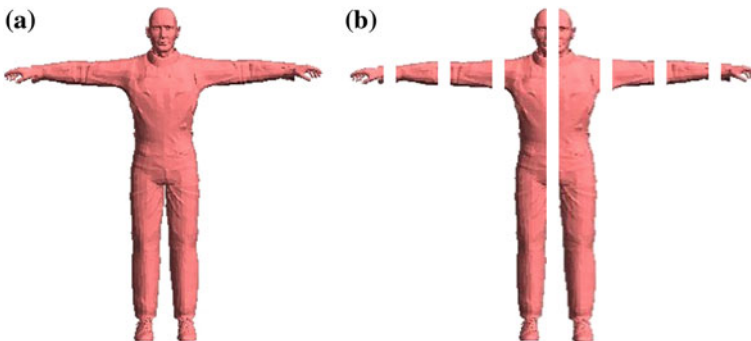


center of the coordinate system and scaled in order to lie within a bounding sphere of radius 1. After translation and scaling, a rotation estimation step based on Continuous Principal Component Analysis (CPCA) is used for rotation estimation in this research, which computes the principal axes of a 3D model based on the continuous triangle set.

Assume that  $N_p$  parts are partitioned by the parallel planes with the XOY, YOZ and XOZ plane respectively. Figure 37.3a is a human 3D model and it is partitioned into 8 parts along its XOY plane, as is shown in Fig. 37.3b. The physical characteristic of each part is computed, and then a feature vector of a whole model is combined.

### 37.2.3 The Inertia Moments Descriptor

For each part of a 3D model, the inertia moment describes the inertia characteristics of the parts with the normalized coordinate axis. Since the part of the model



**Fig. 37.3** A 3D model **a** and its partition along the XOY plane **b**



is represented as the sampled points in this research, the inertia moment of each part is described as follows.

$$I = \sum_{i=1}^N m_i r_i^2 \quad (37.1)$$

where  $i$  is a label assigned to a particular particle,  $m_i$  is the mass of that particle, and  $r_i$  is the shortest distance from the particle's position to the axis of rotation. Assume that there are  $N_i$  randomly sampled points in the part  $i$ . When the 3D model is partitioned along with the XOY plane, the IMD of the sliced parts is denoted as its inertia moment to the Z axis, which is defined as follows.

$$IMD_1 = (I_{11}, I_{12}, \dots, I_{1N_p}), I_{1i} = \sum_{j=1}^{N_i} m_j (x_j^2 + y_j^2), i = 1, \dots, N_p \quad (37.2)$$

Likewise, when a model is partitioned along with the YOZ and XOZ planes, the IMD of the sliced parts to the Y and X axis respectively is shown as Eqs. 37.3 and 37.4.

$$IMD_2 = (I_{21}, I_{22}, \dots, I_{2N_p}), I_{2i} = \sum_{j=1}^{N_i} m_j (x_j^2 + z_j^2), i = 1, \dots, N_p \quad (37.3)$$

$$IMD_3 = (I_{31}, I_{32}, \dots, I_{3N_p}), I_{3i} = \sum_{j=1}^{N_i} m_j (x_j^2 + y_j^2), i = 1, \dots, N_p \quad (37.4)$$

Thus,  $IMD = (IMD_1, IMD_2, IMD_3)$ , and the dimension of  $IMD$  is  $3N_p$ .

### 37.2.4 The Elastic Potential Energy Descriptor

According to Hooke's Law, the restoring force of a spring is proportional to the negative displacement from the equilibrium position of the spring:

$$F = -kx \quad (37.5)$$

where  $F$  is the force that tends to restore a spring to its equilibrium position  $x$  is the displacement of the spring from its equilibrium position, and  $k$  is the spring constant. In physics, this law describes the elastic property of a spring. The corresponding potential energy can be stated as:

$$E = kx^2 \quad (37.6)$$

$$E = \sum_{i=1}^n kx_i^2 \quad (37.7)$$

The equilibrium position is set as the centroid of the 3D model in this research. Thus, the elastic potential energy descriptor (EPED) is given as  $EPED = (E_1, E_2, E_3)$ . The dimension of EPED is  $3N_p$ .

### 37.2.5 The Density Descriptor

The density is a key characteristic to represent the 3D model. The density descriptor (DD) of the sliced part of the 3D model is set as the proportion of the points of the sliced part to the total points, and is shown as the follows.

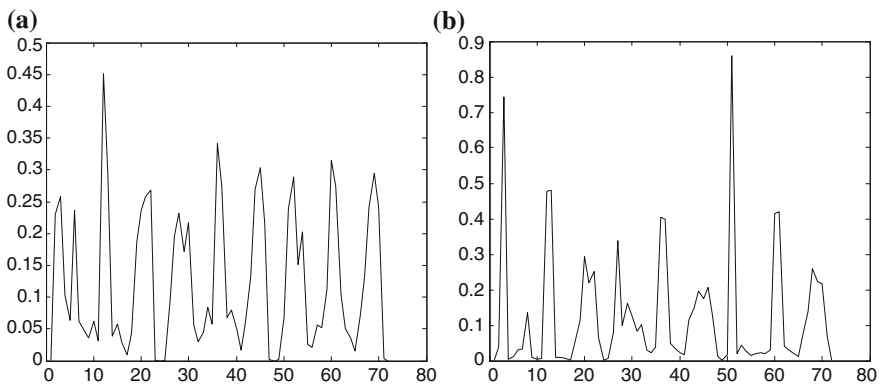
$$DD_i = \frac{N_i}{N} \quad (i = 1 \cdots N_p) \quad (37.8)$$

where  $N$  is the total number of random sampled points of the 3D model, and  $N_i$  is the number of points on the  $N_p$  part. The density descriptor (DD) is given as  $DD = (DD_1, DD_2, DD_3)$ . The dimension of DD is  $3N_p$ .

Thus, the PPD is the synthesized approach of the three physical descriptors mentioned previously, it is expressed as a whole and a single vector as follows:

$$PPD = (IMD, EPED, DD) \quad (37.9)$$

The dimension of the PPD is  $9N_p$ . The PPD of the 3D ant model in Figs. 37.1 and 3D human model in Fig. 37.3 are shown in Fig. 37.4a, b. Obviously, the human and the ant model can be clearly distinguished by using the PPD.



**Fig. 37.4** Partial physical descriptors (PPD) of an ant model in **a** and a human model in **b**

### 37.3 Experimental Results

The database of the Princeton Shape Benchmark (PSB) [9] is used in the experiments. The PSB provides a test data set includes 907 models with 92 classes. The proposed method was tested on a Core2 Quad, 2.7 GHZ system, with 2G of RAM. The precision-recall diagram is used for performance evaluation.

Precision-recall plot: For each query model in class  $C$  and any number  $K$  of top matches, *Recall* is the percentage of models in class  $C$  accurately retrieved within the top  $K$  matches. *Precision* represents the percentage of the top  $K$  matches which are members of class  $C$ . The precision-recall plot indicates the relationship between precision and recall in a ranked list of matches. Curves closer to the upper right corner represent superior retrieval performance.

The D2 [3], AAD [4] and ABD [5] are used in our 3D model retrieval system for testing. The length of D2 descriptor is 64, the dimension of AAD is  $16 \times 8$  and the dimension of ABD is  $16 \times 4 \times 4$ . These descriptors are implemented by us, and normalized distance is used as the distance measure of these descriptors. The Experimental results of D2, AAD, ABD and the proposed PPD are shown in Fig. 37.5.

In Fig. 37.5, obviously, PPD is better than D2, AAD and ABD, furthermore, the dimension of PPD is the shortest one (in the experiment,  $N_p$  is selected as 4, thus, the length of PPD is 36). Thus, PPD obtain good retrieval performance than others.

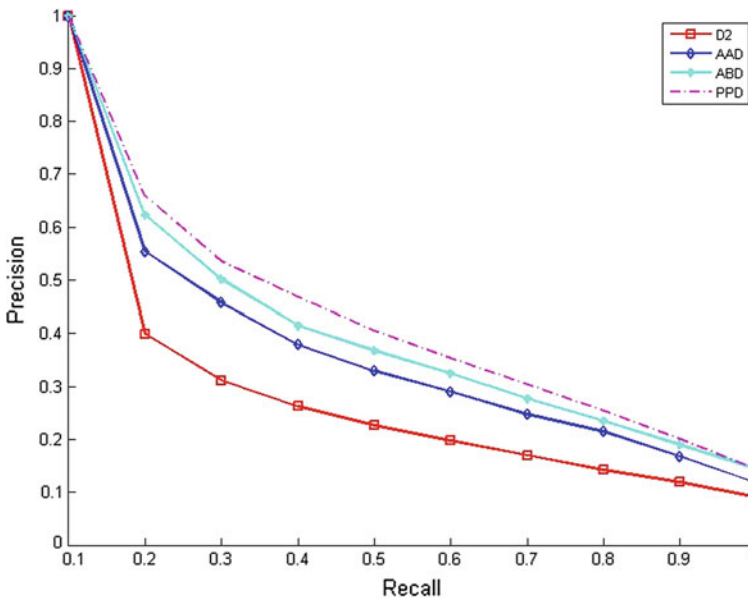


Fig. 37.5 The precision-recall plots of feature fusing for D2 and ABD on the PSB dataset

## 37.4 Conclusion

A novel 3D representation by using sliced parts is proposed. The 3D model is sliced into several parts by the planes paralleling the XOY, YOZ and XOZ plane respectively. The feature of each partial part is extracted. It can be further used for 3D partial searching. The proposed PPD is invariant to the sliced parts, thus, if a 3D model is sliced into few parts, the dimension of PPD is very small. E.g. sliced four parts, the dimension of PPD is 36. PPD is easy to compute, store, and cluster. Further work is to discover spatial relationships among local parts of a model adopt more distinctive physical descriptors and mine associations among these descriptors, in order to support a more effective retrieval.

## References

1. Lin C-H, Huang D-C, Chan Y-K, Chen K-H, Chang Y-J (2011) Fast color-spatial feature based image retrieval methods. *Expert Syst Appl* 38:11412–11420
2. Bustos B, Keim DA, Saupe D, Schreck T, Vranic DV (2005) Feature-based similarity search in 3D object databases. *ACM Comput Surv* 37:345–387
3. Osada R, Funkhouser T, Chazelle B, Dobkin D (2002) Shape distributions. *ACM Trans Graph* 21:807–832
4. Zou KS, Chen ZQ, Ip WH (2011) 3D model retrieval method by using fractal D2 distribution. *Electron Lett* 47(14):796–797
5. Ohbuchi R, Minamitani T, Takei T (2005) Shape-similarity search of 3D models by using enhanced shape functions. *Int J Comput Appl Technol* 23:70–85
6. Reisert M, Burkhardt H (2006) Second order 3D shape features: an exhaustive study. *Comput Graph* 30:197–206
7. Chen D-Y, Tian X-P, Shen Y-T, Ouhyoung M (2003) On visual similarity based 3D model retrieval. *Comput Graph Forum* 22:223–232
8. Vranic DV (2005) Desire: a composite 3D-shape descriptor. In: ICME, IEEE, New York, pp 962–965
9. Shilane P, Min P, Kazhdan M, Funkhouser T (2004) The princeton shape benchmark, In: *International Proceedings of the Shape Modeling*, pp 167–178

# Chapter 38

## Leaf Classification Methods Based on SVM and SIFT

Yida Ye

**Abstract** In this research, Support Vector Machine (SVM) and Scale-Invariant Feature Transform (SIFT) are used to identify plants. For each leaf image, the algorithm localizes the keypoints and assigns orientations for each keypoint. Then it matches the sample leaves with the comparison leaves to find out whether they belong to the same category. After conducting edge detection and feature extraction, the experimental result shows that the method for classification gives average accuracy of approximately 99 % when it is tested on 12 descriptive features.

**Keywords** Leaf classification • SVM • SIFT • Edge detection • Feature extraction

### 38.1 Introduction

Trees can be seen everywhere in our daily life, but seldom do we observe trees carefully. Leaves vary a lot from shape to size and have some interesting specialities such as phototropism [1]. In this paper some features of leaves are first discussed and then several models to classify leaves and weigh leaves are built. Finally, the strength and weakness of our methods and propose some improvements are analyzed.

---

Y. Ye (✉)

School of Electrical and Electronic Engineering, North China Electric Power University,  
2 Beinong Road, Huilongguan Town, Changping District, Beijing, China  
e-mail: sonata1000years@sina.com

## 38.2 Leaf Classification

### 38.2.1 Leaf Classification Based on SVM

#### 38.2.1.1 Feature Extraction

Twelve commonly used digital morphological features, derived from five basic features, are extracted so that a computer or smart phone can obtain feature values quickly and automatically.

#### Basic Geometric Features

- **Diameter:** The diameter is defined as the longest distance between any two points on the margin of the leaf. It is denoted as  $D$ .
- **Physiological length:** The distance between the two terminals of the main veins of the leaf is defined as the physiological length. It is denoted as  $L_p$ .
- **Physiological width:** The maximum length of a line, which is orthogonal to the main vein, is defined as the physiological width. It is denoted as  $W_p$ . Since the coordinates of pixels are discrete, two lines are orthogonal if their degree is  $90 \pm 0.5^\circ$ . The relationship between Physiological Length and Physiological Width is shown in Fig. 38.1.
- **Leaf area:** The value of leaf area is easy to evaluate, just counting the number of pixels of binary value 1 on smoothed leaf image. It is denoted as  $A$ .
- **Leaf perimeter:** Denoted as  $P$ , leaf perimeter is calculated by counting the number of pixels consisting leaf margin.

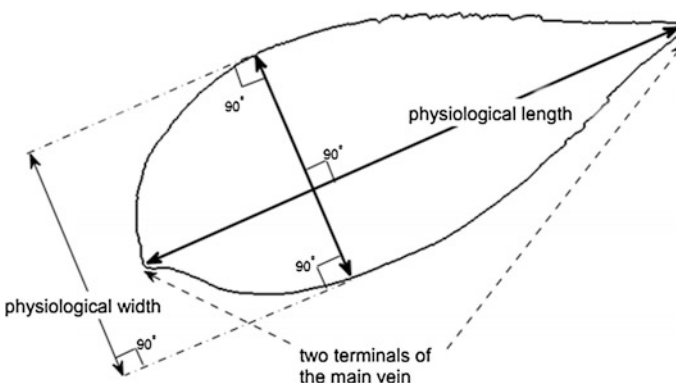


Fig. 38.1 The relationship between  $L_p$  and  $W_p$

### Digital Morphological Features

- **Smooth factor:** smooth factor is defined as the ratio between area of leaf image smoothed by  $5 \times 5$  rectangular averaging filter and the one smoothed by  $2 \times 2$  rectangular averaging filter [2].
- **Aspect ratio:** the aspect ratio is defined as the ratio of physiological length  $L_p$  to physiological width  $W_p$ , thus  $L_p/W_p$ .
- **Form factor:** This feature is used to describe the difference between a leaf and a circle. It is defined as  $4\pi A/P^2$ , where  $A$  is the leaf area and  $P$  is the perimeter of the leaf margin.
- **Rectangularity:** Rectangularity describes the similarity between a leaf and a rectangle. It is defined as  $L_p W_p/A$ , where  $L_p$  is the physiological length,  $W_p$  is the physiological width and  $A$  is the leaf area.
- **Narrow factor:** Narrow factor is defined as the ratio of the diameter  $D$  and physiological length  $L_p$ , thus  $D/L_p$ .
- **Perimeter ratio of diameter:** Ratio of perimeter to diameter, representing the ratio of leaf perimeter  $P$  and leaf diameter  $D$ , is calculated by  $P/D$ .
- **Perimeter ratio of physiological length and physiological width:** This feature is defined as the ratio of leaf perimeter  $P$  and the sum of physiological length  $L_p$  and physiological width  $W_p$ , thus  $P/(L_p + W_p)$ .

#### 38.2.1.2 Support Vector Machine

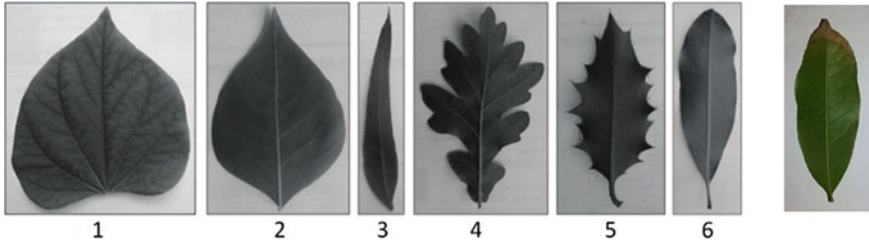
The support vector machine is currently the most popular approach for supervised learning.

SVM finds the OSH (optimum separation hyperplane) by maximizing the margin between the classes [3]. The main concepts of SVM are to first transform input data into a higher dimensional space by means of a kernel function and then construct an OSH between the two classes in the transformed space. Those data vectors nearest to the constructed line in the transformed space are called the support vectors. The SVM estimates a function for classifying data into two classes. Using a nonlinear transformation that depends on a regularization parameter, the input vectors are placed into a high-dimensional feature space, where a linear separation is employed.

To construct a nonlinear support vector classifier, the inner product  $(x,y)$  is replaced by a kernel function  $K(x,y)$

$$f(x) = \text{sgn} \left( \sum_{i=1}^l \alpha_i y_i K(x_i, x) + b \right)$$

where  $f(x)$  determines the membership of  $x$ . In this study, the normal subjects were labeled as  $-1$  and other subjects as  $+1$ . The SVM has two layers. During the



**Fig. 38.2** Leaf class and sample leaf

learning process, the first layer selects the basis  $K(xi, x)$ ,  $i = 1, 2, \dots, N$  from the given set of kernels, while the second layer constructs a linear function in the space. This is equivalent to finding the optimal hyper plane in the corresponding feature space [4, 5]. The SVM algorithm can construct a variety of learning machines using different kernel functions. Based on 100 trials, the result of classification is approximately 99 % accuracy (Fig. 38.2).

## 38.2.2 Leaf Classification Based on SIFT

### 38.2.2.1 SIFT Algorithm

#### Scale-Space Extrema Detection

This is the stage where the interest points, which are called keypoints in the SIFT framework, are detected. For this, the image is convolved with Gaussian filters at different scales, and then the difference of successive Gaussian-blurred images are taken. Keypoints are then taken as maxima/minima of the Difference of Gaussians (DoG) that occur at multiple scales [6]. Specifically, a DoG image  $D(x, y, \sigma)$  is given by

$$D(x, y, \sigma) = L(x, y, k_i\sigma) - L(x, y, k_j\sigma)$$

where  $L(x, y, \sigma)$  is the convolution of the original image  $I(x, y)$  with the Gaussian blur  $G(x, y, \sigma)$  at scale  $k\sigma$ , i.e.,

$$L(x, y, k\sigma) = G(x, y, k\sigma) * I(x, y)$$

#### Keypoint Localization

Scale-space extrema detection produces too many keypoint candidates, some of which are unstable. The next step in the algorithm is to perform a detailed fit to the nearby data for accurate location, scale, and ratio of principal curvatures. This



information allows points to be rejected that have low contrast (and are therefore sensitive to noise) or are poorly localized along an edge.

### Interpolation of Nearby Data for Accurate Position

First, for each candidate keypoint, interpolation of nearby data is used to accurately determine its position. The initial approach was to just locate each keypoint at the location and scale of the candidate keypoint. The new approach calculates the interpolated location of the extremum, which substantially improves matching and stability. The interpolation is done using the quadratic Taylor expansion of the Difference-of-Gaussian scale-space function,  $D(x, y, \sigma)$  with the candidate keypoint as the origin. This Taylor expansion [6] is given by:

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x$$

where  $D$  and its derivatives are evaluated at the candidate keypoint and  $x = (x, y, \sigma)$  is the offset from this point. The location of the extremum,  $\hat{x}$ , is determined by taking the derivative of this function with respect to  $x$  and setting it to zero. If the offset  $\hat{x}$  is larger than 0.5 in any dimension, then that's an indication that the extremum lies closer to another candidate keypoint.

### Eliminating Edge Responses

For poorly defined peaks in the DoG function, the principal curvature across the edge would be much larger than the principal curvature along it. Finding these principal curvatures amounts to solving for the eigenvalues of the second-order Hessian matrix,  $\mathbf{H}$ :

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

The eigenvalues of  $\mathbf{H}$  are proportional to the principal curvatures of  $D$ . It turns out that the ratio of the two eigenvalues, say  $\alpha$  is the larger one, and  $\beta$  the smaller one, with ratio  $r = \alpha/\beta$ , is sufficient for SIFT's purposes. The trace of  $\mathbf{H}$ , i.e.,  $D_{xx} + D_{yy}$ , gives us the sum of the two eigenvalues, while its determinant, i.e.,  $D_{xx}D_{yy} - D_{xy}^2$ , yields the product. The ratio can be shown to be equal to, which depends only on the ratio of the eigenvalues rather than their individual values.

### Orientation Assignment

In this step, each keypoint is assigned one or more orientations based on local image gradient directions. This is key step in achieving invariance to rotation as

the keypoint descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation. The Gaussian-smoothed image  $L(x, y, \sigma)$  at the keypoint's scale  $\sigma$  is taken so that all computations are performed in a scale-invariant manner. For an image sample  $L(x, y)$  at scale  $\sigma$ , the gradient magnitude  $m(x, y)$ , and orientation  $\theta(x, y)$ , are precomputed using pixel differences:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1} \left( \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right)$$

### Keypoint Descriptor

It is important to compute a descriptor vector for each keypoint such that the descriptor is highly distinctive and partially invariant to the remaining variations such as illumination, 3D viewpoint, etc. First a set of orientation histograms are created on  $4 \times 4$  pixel neighborhoods with eight bins each. These histograms are computed from magnitude and orientation values of samples in a  $16 \times 16$  region around the keypoint such that each histogram contains samples from a  $4 \times 4$  subregion of the original neighborhood region. The magnitudes are further weighted by a Gaussian function with  $\sigma$  equal to one half the width of the descriptor window. The descriptor then becomes a vector of all the values of these histograms. Since there are  $4 \times 4 = 16$  histograms each with eight bins the vector has 128 elements. This vector is then normalized to unit length in order to enhance invariance to affine changes in illumination. To reduce the effects of non-linear illumination a threshold of 0.2 is applied and the vector is again normalized (Figs. 38.3, 38.4).

It can clearly draw from both graphs that the heads and tails are matched perfectly and several key points on the margin are also matched [7]. It is difficult to get the optimized match in the absence of a large database on the a leaf shape. However, the effectivity of the SIFT algorithm in leaf classification is quite obvious and can be applied better [8] (Figs. 38.5, 38.6).

**Fig. 38.3** 72 key points

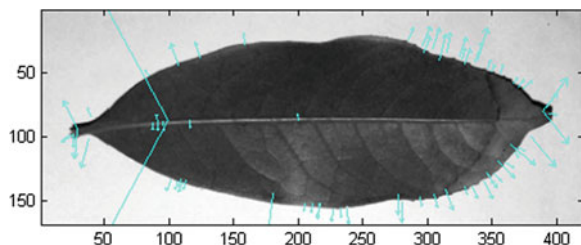


Fig. 38.4 85 key points

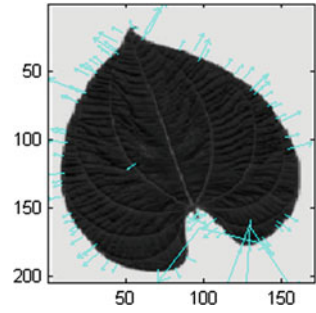


Fig. 38.5 11 matches

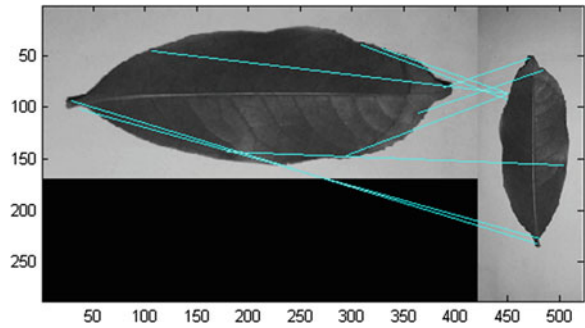
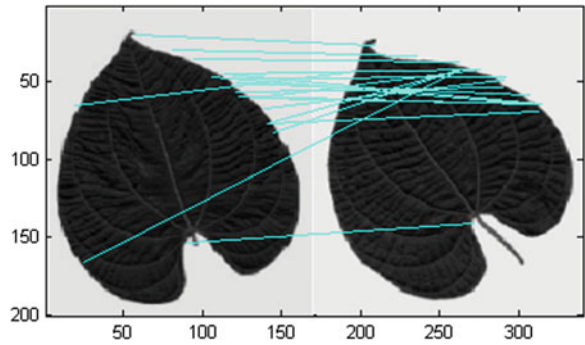


Fig. 38.6 20 matches



### 38.3 Summary

Two models are built to automatically classify leaves. Model one is based on SVM and edge detection is conducted to get a compact and abstract representation of a leaf. Then six typical descriptive features of leaves are applied to train SVM and the accuracy of classification is approximately 99 %. Model two is based on SIFT. The algorithm localizes the keypoints and assigns orientations for each keypoint. The leaves' heads and tails and keypoints along the margins are perfectly matched. The classification will be more accurate if based on a larger comparison group.

## References

1. Kadir A et al (2011) Leaf classification using shape, color, and texture features. *Int J Comput Trends Technol*, July to Aug Issue: 225–230
2. Wu SG et al (2007) A leaf recognition algorithm for plant classification using probabilistic neural network. In: *IEEE international symposium on signal processing and information technology*, IEEE
3. Beghin T, Cope JS, Remagnino P et al (2010) Shape and texture based plant leaf classification. In: *Advanced concepts for intelligent vision systems*, Springer, pp 345–353
4. Singh K, Gupta I, Gupta S (2010) SVM-BDT PNN and fourier moment technique for classification of leaf shape. *Int J Sig Proc Image Proc Pattern Recogn* 3(4)
5. Knight D, Painter J, Potter M, Automatic plant leaf classification for a mobile field guide
6. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2):91–110
7. Tzionas P, Papadakis SE, Manolakis D (2005) Plant leaves classification based on morphological features and a fuzzy surface selection technique. In: *5th international conference technology and automation*, Thessaloniki, Greece, pp 365–370
8. Wang D, Zhang X, Liu Y (2006) Recognition system of leaf images based on neuronal network. *J For Res* 17(3):243–246

# Chapter 39

## Saliency Preserved Image Fusion Using Nonsubsampled Contourlet Transform

Liang Xu, Junping Du, Qingping Li and JangMyung Lee

**Abstract** The visual attention model inspired by the early primate visual system is a very important tool in image processing. Based on the visual attention model, the paper proposes a novel saliency preserved image fusion algorithm with a nonsubsampled contourlet transform (NSCT). The basic idea is that the visual saliency map is first built on the coefficients of the NSCT using the visual attention model, and then is combined with the coefficients of the NSCT to form the activity level which is employed to select the final fused coefficients. The algorithm can transform successfully the visual sensitive information from source images into the fused image which contains abundant detailed contents and preserves effectively the saliency structure while enhances the image contrast. Experiments demonstrate that the proposed algorithm yields the encouraging results.

**Keywords** Saliency map · Nonsubsampled contourlet transform · Image fusion

### 39.1 Introduction

Image fusion technique has huge potential for growth and has been used successfully to many fields, such as remote sensing and medical imaging etc. The image fusion is to combine several source images into a fused image, which contains all important contents from source images and express the more abundant information in a scene. According to the level, image fusion approaches can be

---

L. Xu · J. Du (✉) · Q. Li  
Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia,  
School of Computer Science, Beijing University of Posts and Telecommunications,  
Beijing 100876, China  
e-mail: junpingdu@126.com

J. Lee  
Department of Electronics Engineering, Pusan National University, Busan, Korea

general classified into three types: pixel-level, feature-level or decision-level [1]. Many multi-scale transform (MST) have been developed for fusion. The most popular MST tools are pyramid [2] and wavelet [3] transform. To improve the accuracy of decomposition and reconstruction, the more advanced MST tools are proposed, such as ridgelets [4], contourlets [5] and curvelets [6] etc. In addition, Gemma Piella [7] performs the image fusion by a variational model, and the fused result contains the geometry structure of all the inputs and enhances the contrast for visualization. In [8], a variational approach is proposed based on error estimation theory and partial differential equations for image fusion and denoising.

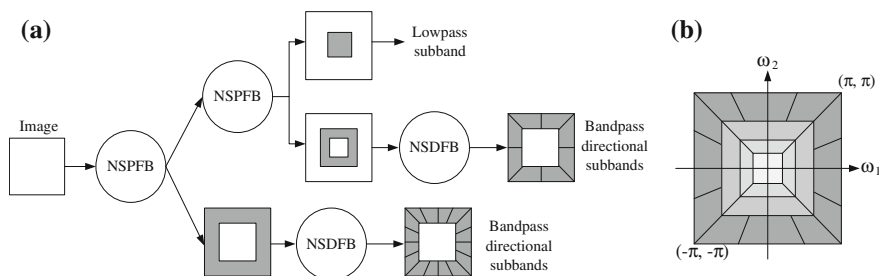
The existing image fusion approaches do not take fully into account the characteristics of HVS, which the human tend to only concern some important areas in a scene. According to the visual perception mechanism, the goal of the proposed algorithm is to preserve the completeness, saliency and sharpness of object areas, and satisfy the requirements of HVS. Thus, based on the NSCT, the paper proposes a novel saliency preserved image fusion algorithm.

### 39.2 Nonsubsampled Contourlet Transform

In this section, we briefly review theory and properties of NSCT, which will be used in the rest of this paper (see [9] for details).

The overcomplete transform NSCT is a shift-invariant version of the contourlet transform and has some excellent properties in process of image decomposition, including shift invariant, multiscale, and multidirection etc. The main components of the NSCT is a nonsubsampled pyramid filter bank (NSPFB) structure for multiscale decomposition and a nonsubsampled directional filter bank (NSDFB) structure for directional decomposition. The NSCT is displayed in Fig. 39.1.

The multiscale property of the NSCT is achieved by using two-channel non-subsampled 2-D filter banks (NSFB), called as NSPFB. The filters for next stage are obtained by upsampling the filters of the previous stage, which the multiscale



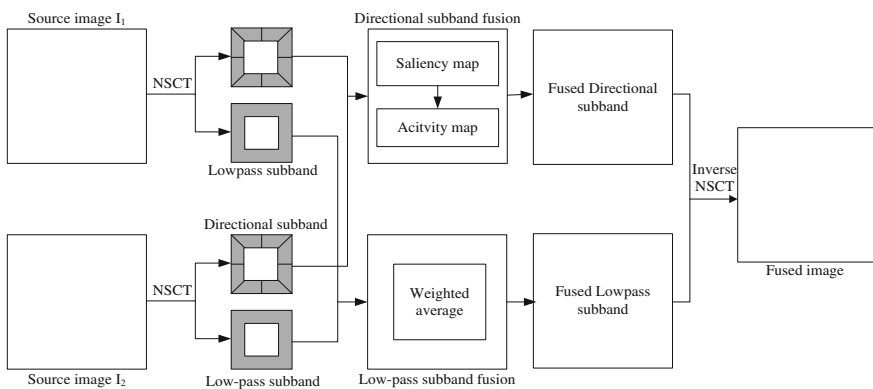
**Fig. 39.1** Nonsubsampled contourlet transform. **a** NSFB structure that implements the NSCT, and **b** idealized frequency partitioning obtained with the proposed structure

property is obtained without the need for additional filter design. The NSDFB, a shift-invariant directional filter bank (DFB) is obtained by eliminating the downsamplers and upsamplers in the DFB. To achieve multi-direction decomposition, the NSDFB is iteratively used. All filter banks in the NSDFB tree structure are obtained from a single NSFb with fan filters. Each filter bank in the NSDFB tree has the same computational complexity as that of the building-block NSFb.

Figure 39.1 shows the NSCT which is constructed by combining the NSPFB and the NSDFB. The two-channel NSFb in the NSPFB and the NSDFB satisfy the Bezout identity and are invertible, so the NSCT is invertible. The key of the NSCT is the filter design problem of the NSPFB and NSDFB. The aim is to design the filters supporting the Bezout identity and obtaining other useful properties. In addition, for a fast implementation, the mapping approach is used to transform the filter into a ladder or lifting structure.

### 39.3 Saliency Preserved Image Fusion

In the section, the proposed saliency preserved image fusion algorithm is presented in detail. The main idea is that the visual saliency map is first built on the coefficients of the NSCT using the visual attention model, and then is combined with the coefficients of the NSCT to form the activity level which is employed to select the final fused coefficients. The fused image has more natural visual appearance and can satisfy the requirements of HVS. The framework of the proposed algorithm is shown in Fig. 39.2. For the clearness of the presentation, we assume that two registered source images are combined.



**Fig. 39.2** Architecture of the proposed algorithm

### 39.3.1 Images Decomposition and Saliency Maps Construction

The decomposition of source images employs the NSCT presented in Sect. 39.2. Input images  $A$  and  $B$  are decomposed into different scale and direction subbands using NSCT. The subband  $\{C_{j_0}^A(x, y), C_{j,l}^A(x, y)\}$  and  $\{C_{j_0}^B(x, y), C_{j,l}^B(x, y)\}$  are obtained, where  $C_{j_0}(x, y)$  denotes the low-pass subband coefficients of the input images at the coarsest scale, and  $C_{j,l}(x, y)$  denotes the high-pass directional subband coefficients at the  $j$ th scale and in the  $l$ th direction.

The saliency maps  $S_{j,l}^A(x, y)$  and  $S_{j,l}^B(x, y)$  are computed on the high-pass directional subbands  $C_{j,l}^A(x, y)$  and  $C_{j,l}^B(x, y)$ , which denotes the  $j$ th scale and  $l$ th direction. The saliency maps provide the selecting index of the coefficients for preserving important information of source images.

Phase spectrum of Fourier transform (PFT) proposed in [10] is employed to a saliency detection model for grayscale image. PFT showed that the saliency map can be easily computed by the phase spectrum of an image's Fourier transform when its amplitude spectrum is at nonzero constant value. Only the phase spectrum is used to reconstruction an image. The reconstruction image has a similar structure with the source image and reflects the saliency information of the source image. The implementation of PFT model consists of three steps. An image is first transformed into frequency domain using Fourier transform, and the amplitude and phase spectrums are then obtained. Finally, the saliency map is obtained by inverse Fourier transform on only the phase spectrum. Given an input image  $I(x, y)$ , three steps have the corresponding Eqs. (39.1–39.3) as follows,

$$F(u, v) = F(I(x, y)) \quad (39.1)$$

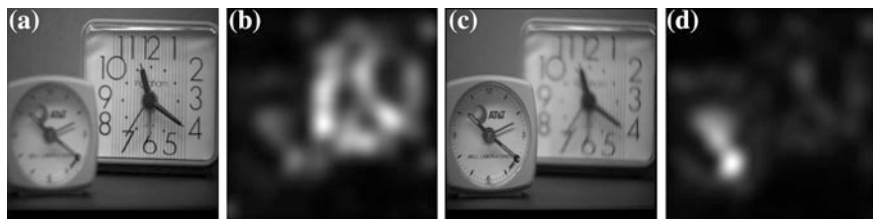
$$P(u, v) = P(F(u, v)) \quad (39.2)$$

$$S(x, y) = g * \|\| F^{-1} \{ \exp^{i \cdot P(u, v)} \} \|\|^2 \quad (39.3)$$

where  $F$  and  $F^{-1}$  denote Fourier transform and inverse Fourier transform.  $P(F)$  is the phase spectrum of  $I$  and  $g$  is a 2D Gaussian filter. The saliency value in location  $(x, y)$  is computed using (39.3).

PFT model is a simple and efficient saliency detection method. An example about the PFT saliency detection is shown in Fig. 39.3. Figure 39.3a, c are two out-of-focus source images which show complementary focus point regions. Figure 39.3b, d are the corresponding saliency maps which indicate different saliency regions of source images. We can observe that the saliency maps present focus point areas in source images. Consequently, the saliency values can be used as the selecting index of subband coefficients.





**Fig. 39.3** The results of saliency detection from two complementary input images. **a, c** Out-of-focus source images, **b, d** Saliency maps from PFT

### 39.3.2 Subband Coefficients Fusion

The high-pass subbands of NSCT decomposition contain abundant detail information and indicate the saliency components of images, e.g. lines, edges and contours etc. In order to preserve the saliency components in the process of image fusion, we propose the saliency preserved fusion rule (SPF) for the high-pass subbands. According to the visual attention mechanism, different regions in an image have varying importance for HVS, so the saliency detection are performed on source images to yield saliency maps which indicate the significance level of every pixel in source images. Based on the characteristic, the saliency maps combined with the coefficients construct the activity maps which denote the energy of coefficients and are used as the selection index of fused coefficients. In addition, the low-pass subband of NSCT decomposition in the coarsest scale contains the main energy of source images, and denotes abundant structural information. The fusion rule of the low-pass subband employs the weighted average of coefficients.

The activity maps of high-pass subbands as the criteria of selecting coefficients are presented as follows. The activity level indicates the magnitude of coefficients. The coefficients of greater energy carry more important information, so the coefficients of greater activity level are selected as the fused coefficients. Now, the activity level in location  $(x, y)$  of high-pass subbands is defined as the product of the coefficient  $C_{j,l}^{A/B}(x, y)$  and the saliency value  $S_{j,l}^{A/B}(x, y)$ , shown as follows.

$$Act_{j,l}(x, y) = C_{j,l}^{A/B}(x, y) \cdot S_{j,l}^{A/B}(x, y) \quad (39.4)$$

The fused coefficients of high-pass subbands denoted as  $F_{j,l}(x, y)$  are defined as,

$$F_{j,l}(x, y) = \begin{cases} C_{j,l}^A(x, y) & \text{if } Act_{j,l}^A(x, y) > Act_{j,l}^B(x, y) \\ C_{j,l}^B(x, y) & \text{otherwise} \end{cases} \quad (39.5)$$

The fused coefficients of low-pass subbands denoted as  $F_{j_0}(x, y)$  are defined as,

$$F_{j_0}(x, y) = 0.5 * C_{j_0}^A(x, y) + 0.5 * C_{j_0}^B(x, y) \quad (39.6)$$

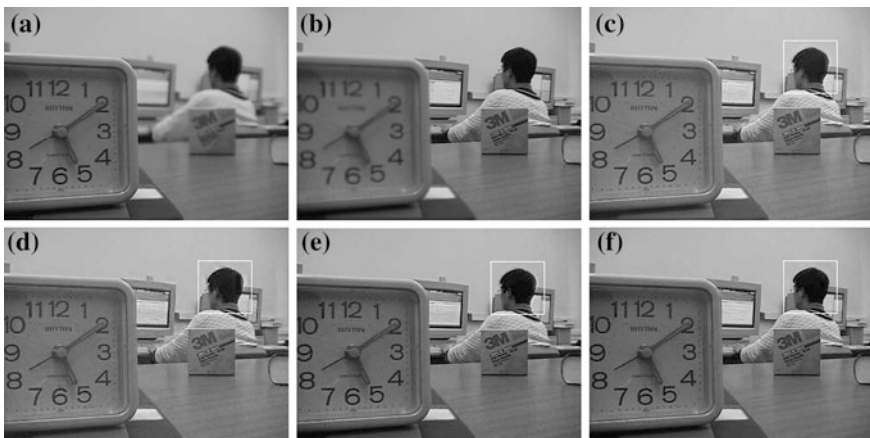
Finally, apply the inverse NSCT to the fused coefficients  $\{F_{j0}(x,y), F_{j,l}(x,y)\}$ , and then obtain the fused image  $F$ .

### 39.4 Experiments and Analysis

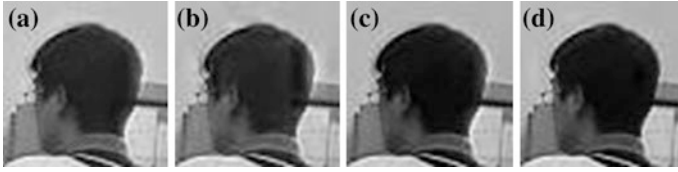
In this section, the proposed NSCT-based saliency preserved fusion (NSCT-SPF) algorithm is tested on several sets of images. For comparison, we use the laplacian pyramid transform (LPT), discrete wavelet transform (DWT), and NSCT-simple. All of these use averaging and absolute maximum selection schemes for merging low- and high-pass subbands. The decomposition level of all of the transforms is three. Three groups of different images were tested to evaluate the performance of the proposed algorithm. It is assumed that source images have been registered. The image data were evaluated using subjective visual inspection and objective assessment tools.

The first set of experiment is two multifocus source images and four fused images shown in Fig. 39.4. The fused images contain all of focus point regions of source images and expand effectively the depth of a scene. The images in Fig. 39.4c–e are not clear enough and have lower contrast; artifacts were also introduced. To observe the image quality in more detail, one area in the fused images was magnified. Figure 39.5a–d shows magnified images of the region marked by the boxes in Fig. 39.4c–f. The fused images in Fig. 39.5a–c have some deformation.

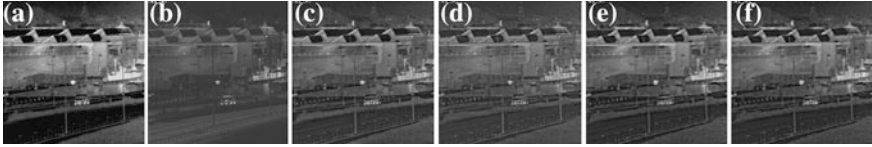
The back of the head in Fig. 39.5a–c appear serious deformation and lead to blur. Figure 39.5d has the better visual quality than others with the best visual effect. Figure 39.6 shows a group of infrared sensor images. A set of spatial out-



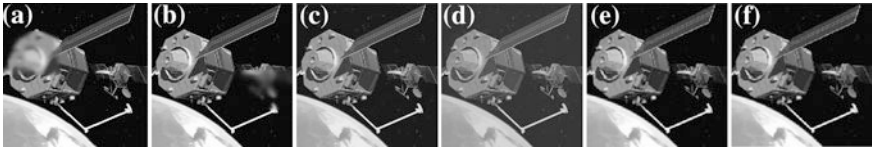
**Fig. 39.4** ‘Lab’ source images and fused images: **a** focus on the left, **b** focus on the right; and fused images using **(c)** LPT, **(d)** DWT, **(e)** NSCT-simple, **(f)** NSCT-SPF methods



**Fig. 39.5** Magnified regions from the fused images in Fig. 39.4(c–f) using **a** LPT, **b** DWT, **c** NSCT-simple, **d** NSCT-SPF methods



**Fig. 39.6** Infrared images fusion (256 level, size of  $252 \times 255$ ) and fused images: **a**, **b** source images; and fused images using, **c** LPT, **d** DWT, **e** NSCT-simple, **f** NSCT-SPF methods



**Fig. 39.7** Spatial source images and fused images: **a** focus on the right satellite, **b** focus on the left satellite; and fused images using, **c** LPT, **d** DWT, **e** NSCT-simple, **f** NSCT-SPF methods

of-focus images are shown in Fig. 39.7, which is obtained by artificial blurring different regions of the ground truth image using a Gaussian filter. Experiment results demonstrate the visual effects of two sets of images in Figs. 39.6 and 39.7 are the same as Fig. 39.4.

In previous discussion, the fusion results of different algorithm have been analyzed by visual aspect. The performance of fusion algorithms need to be further evaluated using objective metric tools. Two metrics are used for evaluation: mutual information (MI) and an objective image fusion performance measure ( $Q_{AB/F}$ ) [11]. The larger the values for the two metrics, the better are the fusion results.

Table 39.1 shows the quality measurement results for fused images in Figs. 39.4, 39.6 and 39.7. Observing the data in Table 39.1, we can see that the LPT and DWT methods are the worst. This is consistent with the subjective visual analysis. Compared with other fusion algorithms, the proposed algorithm NSCT-SPF yields the optimal performance. Experimental results demonstrate that the proposed NSCT-SPF algorithm can preserve the saliency regions of source images, and improve the quality of the fused image.

**Table 39.1** Quality measures for the different fusion methods

Data set	Metric	LPT	DWT	NSCT-simple	NSCT-SPF
Figure 39.4	MI	4.6741	4.4482	4.9683	5.0203
	$Q_{AB/F}$	0.6764	0.6585	0.6989	0.7017
Figure 39.6	MI	2.0389	1.9527	2.2272	2.2391
	$Q_{AB/F}$	0.6064	0.5237	0.6248	0.6460
Figure 39.7	MI	4.9906	4.1951	6.4644	6.4886
	$Q_{AB/F}$	0.8578	0.7332	0.9099	0.9103

## 39.5 Conclusions

The paper proposes a saliency preserved image fusion algorithm based on NSCT. Depending on the human visual attention model, the visual saliency map is first built on the coefficients of the NSCT, and then the algorithm combines the visual saliency map with the coefficients of the NSCT to form the activity level which is employed to select the final fused coefficients. The algorithm can preserve the completeness and the sharpness of object regions. The fused image is more natural and can satisfy the requirement of human visual system. Experiments illustrate that the fusion algorithm improves greatly the quality of the fused images.

**Acknowledgments** This work was supported by the National Basic Research Program of China (973 Program) 2012CB821200 (2012CB821206), the National Natural Science Foundation of China (No. 91024001, No. 61070142) and the Beijing Natural Science Foundation (No. 4111002).

## References

1. Raol JR (2010) Multi-sensor data fusion with matlab. CRC Press Taylor and Francis Group, Boca Raton pp 357–360
2. Petrović VS, Xydeas CS (2004) Gradient-based multiresolution image fusion. *IEEE Trans Image Process* 13(2):228–237
3. Yu XC, Ni F, Long SL, Pei WJ (2012) Remote sensing fusion based on integer wavelet transformation and ordered nonnegative independent component analysis. *Giscience Remote Sens* 49(3):364–377
4. Do MN, Vetterli M (2003) The finite ridgelet transform for image representation. *IEEE Trans Image Process* 12(1):16–28
5. Yang L, Guo BL, Ni W (2008) Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform. *Neurocomputing* 72:203–211
6. Ren S, Cheng J, Li M (2010) Multiresolution fusion of pan and ms images based on the curvelet transform. In: *Proceeding of the Geosci Remote Sens Symp*, 472–475
7. Piella G (2009) Image fusion for enhanced visualization: a variational approach. *Int J Comput Vis* 83(1):1–11
8. Ludusan C, Lavalie O (2012) Multifocus image fusion and denoising: a variational approach. *Pattern Recogn Lett* 33:1388–1396
9. Cunha LD, Zhou JP (2006) The nonsubsampling contourlet transform: theory, design, and applications. *IEEE Trans Image Process* 15(10):3089–3101

10. Guo CL, Ma Q, Zhang LM (2008) Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In: Proceedings of IEEE conference on CVPR, pp 1–8
11. Xydeas C, Petrović V (2000) Objective image fusion performance measure. *Electron Lett* 36(4):308–309

# Chapter 40

## Adaptive Wavelet Packet Filter-Bank Based Acoustic Feature for Speech Emotion Recognition

Yue Li, Guobao Zhang and Yongming Huang

**Abstract** In this paper, a wavelet packet based adaptive filter-bank construction method is proposed, with additive Fisher ratio used as wavelet packet tree pruning criterion. A novel acoustic feature named discriminative band wavelet packet power coefficients (db-WPPC) is proposed and on this basis, a speech emotion recognition system is constructed. Experimental results show that the proposed feature improves emotion recognition performance over the conventional MFCC feature.

**Keywords** Speech emotion recognition · Wavelet packets · Filter-bank design · Speech signal processing

### 40.1 Introduction

With the study on affective computing going deep, automatic emotion recognition is drawing wide attentions from research fields including psychology, linguistics, neuroscience and computer science as an interdisciplinary subject [1]. Automatic emotion recognition from speech shows broad application prospects in fields such

---

Y. Li (✉) · G. Zhang · Y. Huang  
School of Automation, Southeast University, 210096 Nanjing, Jiangsu, China  
e-mail: seuly123@163.com

G. Zhang  
e-mail: guobaozh@seu.edu.cn

Y. Huang  
e-mail: huang\_ym@163.com

G. Zhang  
Key Laboratory of Measurement and Control of Complex Systems of Engineering,  
Ministry of Education, Nanjing, China

as automatic telephone systems [2, 3], interactive movies and online games [4], and intelligent automobile systems [5].

In recent years, wavelet packets (WP) have emerged as an important signal representation and processing scheme [6, 7]. In this paper, the problem of constructing proper tree-structured WP basis that provides efficient discrimination between emotion classes is explored, and a novel WP-based acoustic feature is proposed to capture emotion-discriminating information for speech emotion classification.

The rest of this paper is organized as follows. The proposed discrimination based wavelet packet tree pruning scheme and feature extraction method are presented in Sect. 40.2 together with the realization of speech emotion recognition system. Experimental results are shown and discussed in Sect. 40.3. Finally Sect. 40.4 gives concluding remarks and some ideas about future work.

## 40.2 Feature Extraction and the Proposed System

### 40.2.1 Wavelet Packets and Multi-rate Filter-Bank

As proved in [7], a wavelet packet basis is equivalent with a multi-channel filter-bank followed by a set of aggregated down-samplers. Different admissible WP binary tree structure represents different filter-bank structures, thus providing more flexible frequency partition solutions for signal analysis. For the speech emotion recognition task, our goal is to determine a task-oriented frequency partition solution so that we can locate the emotion information in some specific frequency bands and acquire emotion-related acoustic features by multi-channel filtering of the speech signal. This issue is interpreted as the WP tree pruning problem and will be investigated below.

### 40.2.2 WP Tree Pruning

WP basis can be represented in a binary tree structure where each node has either zero or two children, and we call it an admissible binary tree. Each leaf node in the admissible tree represents a sub-space of the original observation space; different sub-spaces are mutually orthogonal and add up to the original space.

In this subsection, a WP tree pruning approach is proposed referring to [8]. For realization of the tree pruning algorithm, an additive tree pruning criterion named Fisher ratio is described first. On this basis, a bottom-up search is conducted to obtain the optimal binary tree structure.

Before the tree pruning approach is described, some notations are introduced first, following [7] and [9].

Let  $T = \{(0, 0), (1, 0), (1, 1), \dots, (J, 0), \dots, (J, 2^J - 1)\}$  denote an admissible tree with depth  $J \in \mathbb{N}^+$ , where the node  $(j, p)$  is represented by its depth  $(j)$  and the number of nodes on its left at the same depth  $(p)$ . The set of leaf nodes of  $T$  is denoted as  $L(T)$ . Among the sub-trees rooted at  $v_{\text{root}} = (0, 0)$  with  $k$  leaf nodes, the one that maximizes the discrimination measure is denoted by  $T^k$ . The size of  $T$ , denoted by  $|T|$ , is defined as the number of terminal nodes. Let  $\mathbf{x}$  denote a sample in the observation space, [and  $W_{(j,p)}\mathbf{x}$ ] be  $\mathbf{x}$ 's component in the sub-space relating with node  $(j, p) \in T$ . We use measurement  $[M(j, p; \mathbf{x})]$  as the input quantity in discrimination power calculation for  $W_{(j,p)}\mathbf{x}$ . In this paper, the measurement is specified as signal energy, a widely adopted measurement as in [7–9]:

$$M(j, p; \mathbf{x}) = \bar{E}(W_{(j,p)}\mathbf{x}) = \|W_{(j,p)}\mathbf{x}\|^2 / \|\mathbf{x}\|^2 \tag{40.1}$$

where the signal energy is normalized by energy of  $\mathbf{x}$ .

We use  $D(j, p)$  to denote the discrimination measure for node  $(j, p) \in T$  and  $D(T)$  for the whole tree  $T$ . The discrimination measure  $D$  is additive if [8]

$$D(T) = \sum_{(j,p) \in L(T)} D(j, p) \tag{40.2}$$

Given an additive discrimination measure  $D$ , a simple addition is operated instead of computing the functional on the union of the nodes, therefore, a fast algorithm can be applied for the tree-pruning problem. In this paper, Fisher ratio, which has been widely used in feature selection [10], is adopted as tree pruning criterion.

Suppose  $\{\mathbf{x}_i^{(l)}\}_{i=1}^{N_l}, l = 1, 2, \dots, L$  is a training dataset consisting of  $N$  samples, where  $N_l$  denotes the number of signals belonging to class  $l$ . For each node  $(j, p) \in T$ , the Fisher ratio is calculated as follows.

The within-class scatter matrix is defined as

$$S_w(j, p) = \sum_{l=1}^L \frac{N_l}{N} \cdot \Sigma_l(j, p) \tag{40.3}$$

and the between-class scatter matrix is defined by

$$S_b(j, p) = \sum_{l=1}^L \frac{N_l}{N} (\mu(j, p) - \mu_l(j, p))(\mu(j, p) - \mu_l(j, p))^T \tag{40.4}$$

where  $\Sigma_l(j, p)$  and  $\mu(j, p)$  are the covariance matrix and mean value of training samples of the  $l^{\text{th}}$  emotion class, respectively; and  $\mu_l(j, p)$  denotes mean value of the whole set of training samples.

In this paper, with scalar-valued measurement  $[M(j, p; \mathbf{x})]$ , the covariance matrix  $\Sigma_l$  is simplified to the variance of  $[M(j, p; \mathbf{x})]$ . Therefore, both  $S_w(j, p)$  and  $S_b(j, p)$  are scalar-valued, and the Fisher ratio can be written as



$$D_F(j, p) = \frac{S_b(j, p)}{S_w(j, p)} \quad (40.5)$$

Fisher ratio indicates the discrimination power in such a way that  $D_F(\cdot)$  achieves high values when the distribution of data from the same class are concentrated while data of different classes are distributed far from each other.

Equation (40.5) gives the discrimination measure at a single node; and for the whole WP tree, it has been proved that by elaborately selecting the type and order of conjugate mirror filter pair, signal components of different sub-spaces can be well de-correlated so that the overall discrimination measure can be calculated by summing up the values at each leaf node. Therefore, for a given admissible WP tree  $T$ , the overall discrimination power can be calculated in an additive form as in Eq. (40.2).

With the Fisher ratio as tree pruning criterion, the bottom-up tree pruning algorithm proposed in [8] and [11] can be conducted for WP tree pruning, and a sequence of  $\{T^k\}$  with  $k$  ( $1 \leq k \leq |T|$ ) leaf nodes can be constructed.

### 40.2.3 Discriminative Band Wavelet Packet Power Coefficients (db-WPPC)

In the emotion recognition task we intend to catch emotion information embedded in the speech signal, while other irrelevant information such as speaker identity and speech content can be omitted. Guided by this thought, a novel feature is proposed in this paper, which we can refer to as the discriminative band wavelet packet power coefficients (db-WPPC). The block diagram of db-WPPC feature extraction is shown in Fig. 40.1.

Before feature extraction, the speech signal is first pre-emphasized, blocked into 32 ms frames with 16 ms overlap between adjacent frames and then multiplied by Hamming window. By selecting a WP tree structure  $T^k$ , a filter-bank structure with  $k$  sub-bands is constructed. And then a set of first 12 most discriminative sub-bands is selected from the original WP filter-bank using a ranking scheme as proposed in [8] to preserve the sub-bands where the most discriminative emotion information is located. The speech signal is filtered by the selected set of sub-band filters, followed by log-energy calculation. The log-energy of the frame is also calculated to form the feature vector. Finally, the delta and acceleration coefficients are appended to the feature vector and the 39-dimensional db-WPPC feature is derived. In this paper, the Fisher ratio used as discrimination

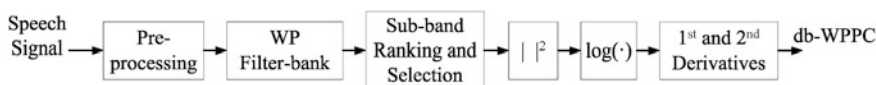


Fig. 40.1 Db-WPPC feature extraction

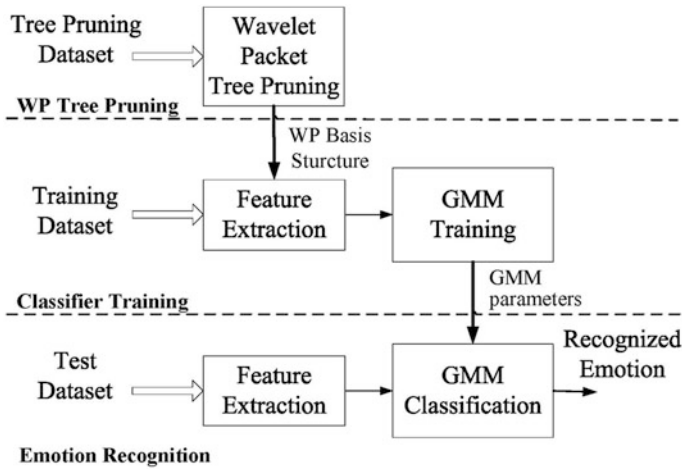


Fig. 40.2 Block diagram of the proposed system

measure in the WP tree pruning step is directly available for the sub-band ranking scheme, since it provides good information of discrimination power of the frequency band.

#### 40.2.4 Proposed System

The framework of proposed speech emotion recognition system is shown in Fig. 40.2. According to the block diagram, we can divide the speech emotion recognition task into three blocks: the WP tree pruning, the classifier training and the emotion recognition; and the whole emotional speech database is divided into the tree-pruning dataset, the training dataset and the test dataset accordingly. Gaussian mixture model (GMM) is adopted as classifier in this paper. For each emotion class a GMM is trained with the training dataset.

### 40.3 Experiments

#### 40.3.1 Experimental Setup

Experiments were conducted on the Berlin emotional speech database [12], which contains seven simulated emotions (anger, boredom, disgust, fear, joy, neutral and sadness). Ten German sentences (five short and five longer) that contain no emotional bias were selected as text material and a total of 535 utterances were

produced by 10 German actors (five female and five male). In this paper, six emotions (no disgust) with a sum of 489 utterances were used for the classification task.

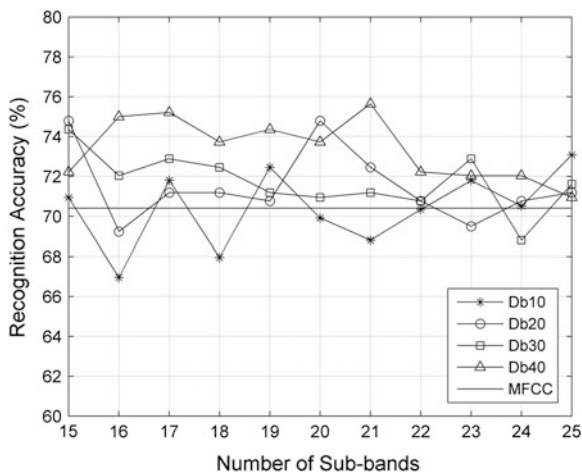
Twenty percent of the utterances were randomly selected to form the tree pruning dataset, and we applied 5-fold cross validation on the remaining 80 % utterances. The maximum WP decomposition level was set to  $J = 5$ ; and the Daubechies wavelets of order 10, 20, 30 and 40 were chosen for wavelet packet decomposition. The number of sub-bands in the filter-bank varied in the range of 15–25. The number of Gaussian mixture components was set to 16 for each GMM.

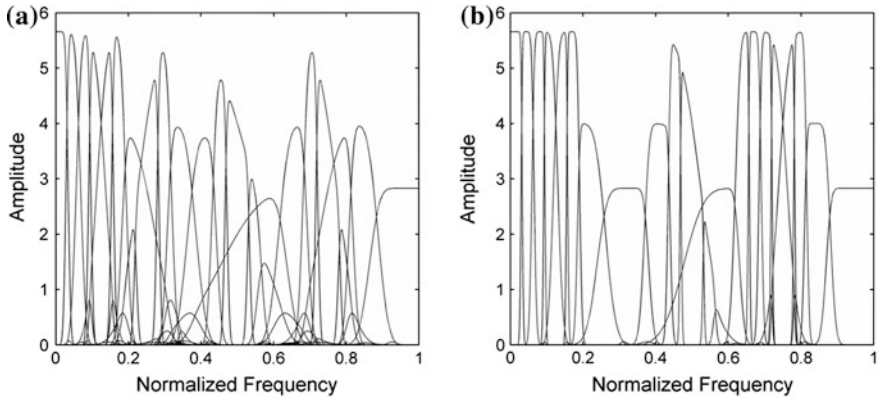
As a benchmark, the conventional MFCC feature was calculated by passing the speech frame through a set of 20 triangular band-pass filters that are equally spaced along the Mel frequency, and then applying Discrete Cosine Transform on the sub-band log-energy coefficients. The first 12 Cepstral coefficients were adopted together with the log-energy of the frame. The first and second order derivatives were appended to the feature vector to form the 39-dimensional MFCC feature.

### 40.3.2 Experimental Results of the Proposed System

The experimental results are illustrated in Fig. 40.3. From Fig. 40.3, it can be seen that with the order of Daubechies filters increases, better classification performance is achieved. This is intuitive since when the filter order increases, better frequency selection property is gained, and thus the additive property of tree pruning criterion can be better satisfied. To be more specific, let us consider the frequency responses of 20-channel filter-banks generated using Daubechies filters of order 10 and 40 illustrated in Fig. 40.4. From Fig. 40.4, it is noticeable that with the higher order adopted, less frequency aliasing is observed between different sub-bands.

**Fig. 40.3** Six-class emotion classification accuracies of GMM classifiers with different Daubechies filter orders as a function of sub-band number





**Fig. 40.4** Frequency responses of the 20-channel wavelet packet filter-banks **a** Daubechies 10, **b** Daubechies 40

**Table 40.1** The 12 most discriminative frequency bands ranked by Fisher score, obtained from the 21-channel filter-bank structure of Daubechies 40

Rank	Frequency interval	Score	Rank	Frequency interval	Score
1	0–250 Hz	0.253	7	2500–3000 Hz	0.031
2	250–500 Hz	0.062	8	2250–2500 Hz	0.028
3	2000–2250 Hz	0.059	9	4000–4500 Hz	0.022
4	3000–4000 Hz	0.046	10	1000–1500 Hz	0.019
5	1750–2000 Hz	0.043	11	1500–1750 Hz	0.017
6	500–750 Hz	0.042	12	4500–5000 Hz	0.013

A highest classification accuracy of 75.64 % is achieved with Daubechies filter of order 40, 21-channel filter-bank. This result outperforms the conventional MFCC feature, a benchmark of 70.41 %. The corresponding discriminative sub-bands are listed in Table 40.1, which gives us an insight into the emotion-related information distribution among the frequency bands.

### 40.4 Conclusion

In this paper we explored a wavelet packet based acoustic feature extraction approach for speech emotion recognition. Tree pruning algorithm was applied with Fisher ratio proposed as tree pruning criterion to adapt the WP filter-bank structure to the decision task. And on this basis, a novel short-time acoustic feature named discriminative band wavelet packet power coefficients (db-WPPC) was proposed for the emotion classification task. Speech emotion recognition system was built and experiments were carried out on the Berlin emotional speech database to evaluate the proposed feature extraction scheme. Experimental results demonstrate the superiority of the proposed feature sets over conventional MFCC feature.

Future work includes investigating more effective WP tree pruning methods and seeking for robust feature representation for speech signal, as well as developing efficient classification techniques for automatic speech emotion recognition.

## References

1. Zeng ZH, Pantic M, Roisman GI et al (2009) A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Trans Pattern Anal Mach Intell* 31:39–58
2. Morrison D, Wang RL, De Silva LC (2007) Ensemble methods for spoken emotion recognition in call-centres. *Speech Commun* 49:98–112
3. Petrushin V (2000) Emotion recognition in speech signal: experimental study, development, and application. Presented at the ICSLP2000, Beijing
4. Caponetti L, Buscicchio CA, Castellano G (2011) Biologically inspired emotion recognition from speech. *EURASIP J Adv Sign Process* 2011:24
5. Malta L, Miyajima C, Kitaoka N et al (2009) Multimodal estimation of a driver's spontaneous irritation, In: *Intelligent Vehicles Symposium, IEEE*, pp 573–577
6. Daubechies I (1992) *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia
7. Pavez E, Silva JF (2012) Analysis and design of wavelet-packet cepstral coefficients for automatic speech recognition. *Speech Commun* 54:814–835
8. Saito N, Coifman RR (1994) Local discriminant bases, In: *Proceedings of SPIE 2303, Mathematical Imaging: Wavelet Applications in Signal and Image Processing*, pp 2–14
9. Silva J, Narayanan SS (2009) Discriminative wavelet packet filter bank selection for pattern recognition. *IEEE Trans Sign Process* 57:1796–1810
10. Wu SQ, Falk TH, Chan WY (2011) Automatic speech emotion recognition using modulation spectral features. *Speech Commun* 53:768–785
11. Scott C (2005) Tree pruning with subadditive penalties. *IEEE Trans Sign Process* 53:4518–4525
12. Burkhardt F, Paeschke A, Rolfes M et al (2005) A database of german emotional speech, In: *Proceeding INTERSPEECH, ISCA, Lisbon, Portugal*, pp 1517–1520

# Chapter 41

## A Novel Decision-Based Algorithm for Removal of Highly Corrupted Images

Yiyan Wang, Zhuoer Wang and Di Zhou

**Abstract** The major drawback of recent image filtering algorithms is lack of the ability of removing high density salt-and-pepper noise. To alleviate this limitation, an improved decision-based algorithm is proposed. Firstly, according to the characteristics of salt-and-pepper noise and local gray-scale feature of pixels, this algorithm separates noise pixels and signal pixels. Then the noise pixels are recovered by the median value of the neighboring noise-free pixel values, while the signal pixels hold their gray values without changing. Different gray-scale and color images have been tested by using the proposed algorithm (PA), simulation results show that this method has the better ability of removing noises and preserving the partial details of images in comparison with some recent methods especially when the noise density is very high.

**Keywords** Image denoising · Salt-and-pepper noise · Noise detection · Median filter

---

Y. Wang (✉)

Department of Physics and Engineering Technology, Sichuan University of Arts and Science, Dazhou Sichuan 635000, China  
e-mail: wang\_yiyan8207@163.com

Y. Wang

Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, Nanjing Jiangsu 210096, China

Z. Wang

Library of Sichuan University of Arts and Science, Dazhou Sichuan 635000, China  
e-mail: vivianwang8403@163.com

D. Zhou

Department of Science and Technology, Sichuan University of Arts and Science, Dazhou Sichuan 635000, China  
e-mail: sion2005@163.com

## 41.1 Introduction

Images are often corrupted by impulse noise during acquisition and transmission; thus, an efficient noise suppression technique is required before subsequent image processing operations [1]. Based on the noises values, the noise can be classified as the fixed-values impulse noise, also named salt-and-pepper noise; and the more difficult random valued impulse noise [2]. In early development of image processing linear filters were the primary tools. But linear filters have poor performance in the presence of noise that is additive in nature. In image processing linear filters tend to blur the edges and do not remove impulse noise effectively. Non-linear filters are developed to overcome these limitations. Median filter [3] is widely used in impulse noise removal methods due to its effective noise suppression capability and high computational efficiency. The main drawback of a standard median filter (SMF) is that it is effective only for low noise densities. At high noise densities, SMFs often exhibit blurring for large window sizes and insufficient noise suppression for small window sizes. However, most of the median filters operate uniformly across the image and thus tend to modify both noise and signal pixels. Consequently, the effective removal of impulse often leads to images with blurred and distorted features. The ideal approach is to apply the filtering technique only to noisy pixels, without changing the uncorrupted pixel values. Nonlinear filters such as Adaptive Median Filter (AMF), switching-based median filter [4–6] can be used for discriminating possible corrupted and uncorrupted pixels, then the noisy pixels will be replaced by using median value or its variant while the uncorrupted pixels will be left unchanged. The performance of AMF is good at lower noise density levels, due to the fact that there are only fewer corrupted pixels that are replaced by the median values. At higher noise densities, window size has to be increased to get better noise removal which will lead to less correlation between corrupted pixel values and replaced median pixel values. In decision-based or switching median filter the decision is based on a pre-defined threshold value. The main drawback of this method is that defining a robust decision measure is difficult. Also the noisy pixels are replaced by some median value in their vicinity without taking into account local features such as the possible presence of edges. Hence details and edges are not recovered satisfactorily, especially when the noise level is high. Recently, Decision Based Median Filtering Algorithm (DBA) [7] was proposed to remove high density salt and pepper noise, which replaces the corrupted pixels by the median or the neighboring pixels value by using a fixed window size of  $3 \times 3$ . Although the recent method [7] showed promising results, at higher noise densities the median may also be a noise pixel and this produces streaking at higher noise densities. To overcome this problem, Nair and Kevathy [8] have proposed an improved algorithm. The corrupted pixels can either be replaced by the median pixel or, the mean of the neighborhood processed pixels. Here, we suggest a new method for corrupted pixels are recovered only by the median value of the neighboring noise-free pixel values. We do not use the gray-scale information of noise pixel itself to remove noise because

the gray value of salt-and-pepper noise is not related to the original pixel. Meanwhile by transforming the noise pixels into noise-free pixels we could avoid the noise spreading in the neighborhood. In addition, our proposed algorithm (PA) also uses fixed length window size of  $3 \times 3$ . The experimental results show that the proposed algorithm has better Peak Signal-to-Noise Ratio (PSNR) and Mean Structured Similarity Index (MSSIM) values in comparison with some recent methods when the noise density is very high ( $> 50\%$ ).

## 41.2 Proposed Algorithm

### 41.2.1 Noise Model

In this approach noise is modeled as salt-and-pepper impulse noise as practiced [9]. Pixels are randomly corrupted by two fixed extremal values, 0 and 255 (for 8-bit monochrome image), generated with the same probability. That is, for each image pixel at location  $(i, j)$  with intensity value  $s_{i,j}$ , the corresponding pixels of the noisy image will be  $x_{i,j}$ , in which the probability density function of  $x_{i,j}$  is

$$f(x) = \begin{cases} \rho/2, & \text{for } x = 0 \\ 1 - \rho, & \text{for } x = s_{i,j} \\ \rho/2, & \text{for } x = 255 \end{cases} \quad (41.1)$$

where  $\rho$  is the noise density.

### 41.2.2 Noise Detection and Removal

In SMF, every pixel is processed and is replaced by the median of its neighborhood values. In our proposed algorithm (PA) salt-and-pepper noises are first detected based on the minimum(0) and maximum(255) value. If the pixel being currently processed lies inside the dynamic range [0, 255] then it is considered as signal pixel and no modification is made to that pixel. Otherwise it is considered as a noisy pixel and will be replaced by the median value of the neighboring noise-free pixel values.

$$\begin{bmatrix} S(i-1, j-1) & S(i-1, j) & S(i-1, j+1) \\ S(i, j-1) & S(i, j) & S(i, j+1) \\ S(i+1, j-1) & S(i+1, j) & S(i+1, j+1) \end{bmatrix} \quad (41.2)$$

In the  $3 \times 3$  window above,  $S(i-1, j-1)$ ,  $S(i-1, j)$ ,  $S(i-1, j+1)$  and  $S(i, j-1)$  indicates already processed pixel values,  $S(i, j)$  indicates the current pixel being processed, and  $S(i, j+1)$ ,  $S(i+1, j-1)$ ,  $S(i+1, j)$  and  $S(i+1, j+1)$  indicates the pixels yet to be processed.



The steps of the PA are elucidated as follows:

**Step 1:** Select a two dimensional window  $W$  of size  $3 \times 3$ . Assume that the pixel being processing is  $S(i, j)$ .

**Step 2:** If  $0 < S(i, j) < 255$ , the  $S(i, j)$  is signal pixel and its values is left unchanged. Otherwise  $S(i, j)$  is a noisy pixel.

**Step 3:** Select all noise-free pixels of window  $W$  as  $W^o$ . Then compute  $W_{med}^0$ , the median of the pixel values in the window  $W^o$ . Obviously,  $W^o$  includes the already processed pixel values such as  $S(i-1, j-1)$ ,  $S(i-1, j)$ ,  $S(i-1, j+1)$  and  $S(i, j-1)$ .

**Step 4:** If  $S(i, j)$  is a noisy pixel, it will be replaced by the  $W_{med}^0$ .

**Step 5:** Repeat Step 1–4 until all the pixels in the entire image are processed.

In the PA, we do not use the gray-scale information of noise pixel itself to remove noise because the gray value of salt-and-pepper noise is not related to the original pixel. Meanwhile by transforming the noise pixels into noise-free pixels we could avoid the noise spreading in the neighborhood.

### 41.2.3 Color Image Denoising

At the most directly used color space for digital image processing, the RGB color space is chosen in our work to represent the color images. In the RGB color space, each pixel at the location  $(i, j)$  can be represented as color vector  $P_{i,j} = (P_{i,j}^R, P_{i,j}^G, P_{i,j}^B)$ , where  $P_{i,j}^R$ ,  $P_{i,j}^G$  and  $P_{i,j}^B$  are the read(R), green(G), and blue(B) components, respectively. The noisy color images are modeled by injecting the salt-and-pepper noise randomly and independently to each of these color components. That is, when a color image is being corrupted by the noise density  $\rho$ , it means that each color component is being corrupted by  $\rho$ . Thus, for each pixel  $P_{i,j}$ , the corresponding pixel of the noise image will be denoted as  $X_{i,j} = (X_{i,j}^R, X_{i,j}^G, X_{i,j}^B)$ , in which the probability density function of each color components can be the same with the noise model of type (41.1).

The process of extending the noise detection algorithm to corrupted color images is straightforward. PA will be simply applied to R-, G-, and B-planes individually, and then combined to form the restored color image.

## 41.3 Simulation Results

The performance of the algorithms are tested using  $512 \times 512$  images such as “man”, “baboon”, “lena”, “peppers” (See Fig. 41.1), and with their dynamic range of values[0, 255]. In the simulation, images will be corrupted by salt-and-pepper noise at different level of noise densities, and the restoration performance are quantitatively measured by Peak Signal-to-Noise Ratio.

(PSNR) and Mean Structured Similarity Index (MSSIM) defined as follows.

$$PSNR = 10 \log 10 \left( \frac{(M \times N) \max(x_{i,j})^2}{\sum_{i,j} (x_{i,j} - y_{i,j})^2} \right) \quad (41.3)$$

$$MSSIM(X, Y) = \frac{1}{M} \sum_{j=1}^M SSIM(x_j, y_j) \quad (41.4)$$

where  $M$  and  $N$  are the total number of pixels in the horizontal and the vertical dimensions of the image;  $x_{i,j}$ ,  $n_{i,j}$  and  $y_{i,j}$  denote the original, corrupted and distorted image pixels, respectively. In (41.4),  $X$  and  $Y$  are the original and the distorted images, respectively;  $x_j$  and  $y_j$  are the image contents at the  $j$ th local window;  $M$  is the number of local windows of image; and SSIM is defined as follow [10]:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (41.5)$$

where  $\mu_x$  is the average of  $x$ ;  $\mu_y$  is the average of  $y$ ;  $\sigma_x^2$  is the variance of  $x$ ;  $\sigma_y^2$  is the variance of  $y$ ;  $\sigma_{xy}$  is the covariance of  $x$  and  $y$ ;  $c_1 = (k_1L)^2$ ,  $c_2 = (k_2L)^2$  are two variables to stabilize the division with weak denominator;  $L$  is the dynamic range of the pixel-values ( $L = 255$ ); and  $k_1, k_2 < 1$  ( $k_1 = 0.01, k_2 = 0.03$ ).

Figures 41.2 and 41.3 and Table 41.2 show the comparison of different filters, performed on gray-scale image “man” and “baboon” at different noise densities, respectively, in terms of MAE, PSNR and MSSIM. Depending on the noise density, window size varies from  $3 \times 3$  to  $15 \times 15$  for SMF[3] and AMF[4], to yield better result. For all level of noise densities DBA [7], DBAM [5] and PA use fixed size window of  $3 \times 3$  (See Table 41.1), all the filters are implemented in MATLAB 7.0 on a PC equipped with 2.67 GHz CPU and 1 GB RAM memory. To make a reliable comparison, each method is run 10 times in every noise density and the result is obtained by averaging over all experiments.

The results of PA (shaded in the Table 41.2 and Fig. 41.3) indicate better PSNR and MSSIM values compared with various filters, namely, SMF[3], AMF[4],

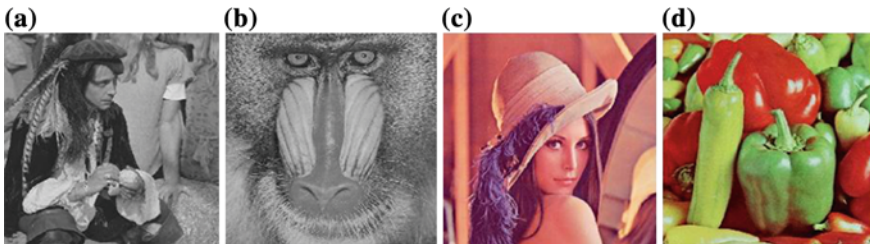
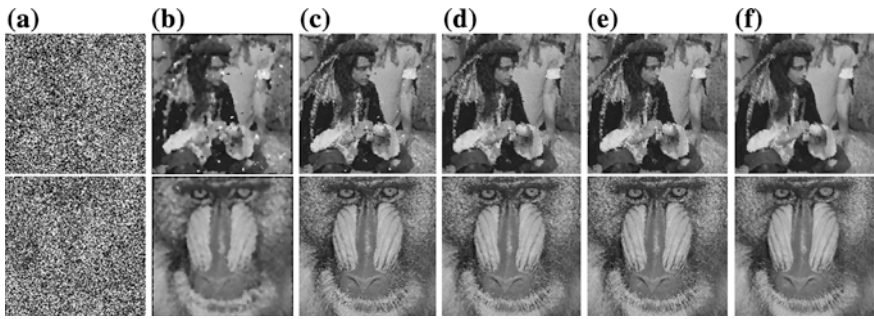
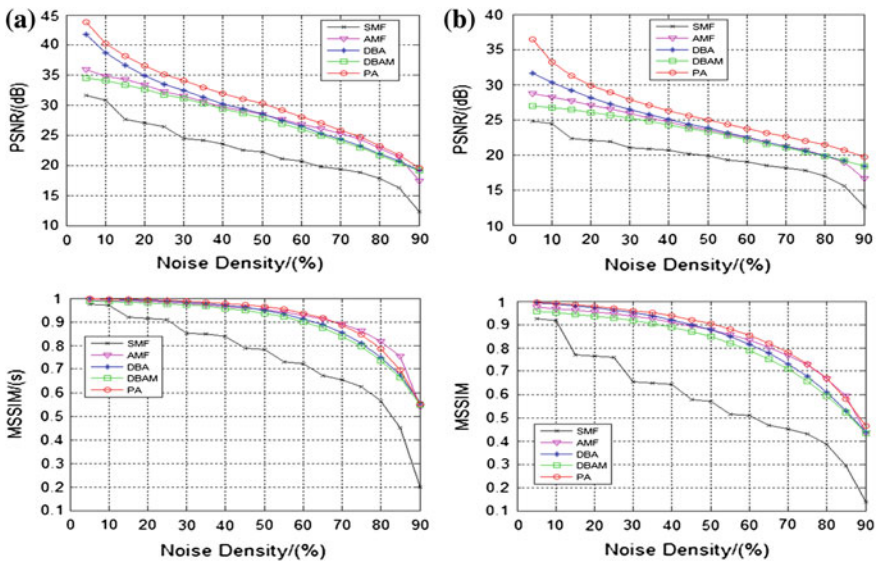


Fig. 41.1 Test images: a man, b baboon, c lena, d peppers



**Fig. 41.2** Simulation results of different filters column. **a** Noise corrupted image. **b** Output for SMF. **c** Output for AMF. **d** Output for DBA. **e** Output for DBAM. **f** Output for PA. Row 1 shows the “man” image corrupted by 80 % noise. Row 2 shows the “baboon” image corrupted by 70 % noise



**Fig. 41.3** Comparison graph for quantitative parameters of various filters for different noise densities column. **a** Results for “man” image. **b** Results for “baboon” image. Row 1 shows PSNR versus noise density. Row 2 shows MSSIM versus noise density

DBA[7], and DBAM[5]. Meanwhile, the PA also can be used for color image. The performance of the color image restoration process is also quantified using PSNR and MSSIM (See Figs. 41.4 and 41.5 and Table 41.3). At a result of this, the PA removes the noise effectively even at noise level as high as 90 % and preserves the edges without any loss up to 80 % of the noise level.

**Table 41.1** Suggested window size for the noise density level

Noise Density	Method				
	SMF	AMF	DBA	DBAM	PA
$0 < p \leq 0.15$	$3 \times 3$	$3 \times 3$	$3 \times 3$ (fixed)	$3 \times 3$ (fixed)	$3 \times 3$ (fixed)
$0.15 < p \leq 0.3$	$5 \times 5$	$5 \times 5$			
$0.3 < p \leq 0.45$	$7 \times 7$	$7 \times 7$			
$0.45 < p \leq 0.6$	$9 \times 9$	$9 \times 9$			
$0.6 < p \leq 0.7$	$11 \times 11$	$11 \times 11$			
$0.7 < p \leq 0.8$	$13 \times 13$	$13 \times 13$			
$0.8 < p \leq 0.9$	$15 \times 15$	$15 \times 15$			

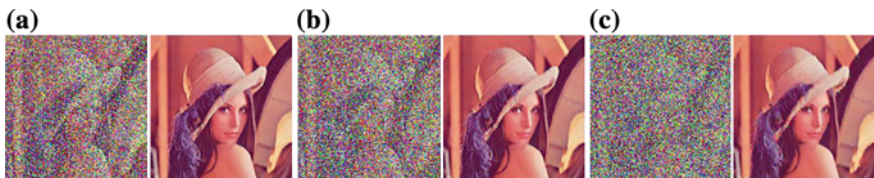
**Table 41.2** PSNR and MSSIM for various filters for “man” and “baboon”(gray images) at various highly noise densities

Test images	Quantitative parameters	SMF	AMF	DBA	DBAM	PA
Man (80 %)	PSNR	17.7800	22.8459	21.9517	21.6742	23.2539
	MSSIM	0.5645	0.8165	0.7487	0.7345	0.7848
Baboon (70 %)	PSNR	18.0720	21.2136	21.2257	21.0425	22.5899
	MSSIM	0.4523	0.7668	0.7309	0.7091	0.7811

The better values of PSNR and MSSIM are signed with black font

**Table 41.3** PSNR and MSSIM for proposed algorithm(PA) for “lena” and “pepper”(color images) at different highly noise densities

Quantitative parameters	lena			peppers		
	60 %	70 %	80 %	70 %	80 %	90 %
PSNR	29.7458	27.6198	24.6768	24.3158	22.0935	18.1855
MSSIM	0.9490	0.9156	0.8435	0.9043	0.8291	0.6309



**Fig. 41.4** Corrupted and denoised images of “lena” at different highly noise densities. **a** 60 % noise density. **b** 70 % noise density. **c** 80 % noise density



**Fig. 41.5** Corrupted and denoised images of “peppers” at different highly noise densities. **a** 70 % noise density. **b** 80 % noise density. **c** 90 % noise density

## 41.4 Conclusion

An efficient non-linear algorithm to remove high density salt-and-pepper noise is proposed. We propose a procedure for noise value detection using the characteristics of salt-and-pepper noise and local gray-scale feature of pixels. For image restoration, the corrupted pixels are recovered only by the median value of the neighboring noise-free pixel values. We do not use the gray-scale information of noise pixel itself to remove noise because the gray value of salt-and-pepper noise is not related to the original pixel. Meanwhile by transforming the noise pixels into noise-free pixels we could avoid the noise spreading in the neighborhood. In addition, the PA uses fixed length size  $3 \times 3$  window having only neighbors of the corrupted pixel that have higher correlation, this provides more edge details, leading to better edge preservation.

The performance of the algorithm has been tested across a wide range of noise densities varying from 5 to 90 % for gray-scale images. Results reveal that the restored images that are obtained by the PA have better objective quality and subjective vision effect in comparison with other existing algorithms. Meanwhile, the PA also can be used for color image. Even at high noise density levels, the PA can suppress noise very effectively and preserve image details very well.

**Acknowledgments** This work was supported in part by the Open Scientific Research Project of Artificial Intelligence Key Laboratory of Sichuan Province, China under Grant No. 2010RY009, by Scientific Research Fund of Sichuan Provincial Education Department, China under Grant No. 13ZB0098 and No. 10ZB135, and by a Grant from the Key Programs of Sichuan University of Arts and Science, China under Grant No. 2012Z002Z, which are gratefully acknowledged.

## References

1. Bovik A (2000) Handbook of Image and Video Processing. Academic Press, New York, pp 461–474
2. Pitas I, Venetsanopoulos AN (1990) Nonlinear digital filters: principles and application. Kluwer Academic, Boston
3. Gallagher Jr NC, Wise GL (1981) A theoretical analysis of properties of the median filters. IEEE Trans Acoust Speech, Signal Process 29(1):1136–1141
4. Hwang H, Haddad RA (1995) Adaptive median filters: new algorithms and results. IEEE Trans Image Process 4(4):499–502
5. Wang Z, Zhang D (1999) Progressive switching median filter for the removal of impulse noise from highly corrupted images. IEEE Trans Circ Syst-II: Analog and Digital Sign Process 46(1):78–80
6. Eng HL, Ma KK (2001) Noise adaptive soft-switching median filter. IEEE Trans Image Process 10(2):242–251
7. Srinivasan KS, Ebenezer D (2007) A New fast and efficient decision-based algorithm for removal of high-density impulse noises. IEEE Sign Process Lett 14(3):189–192
8. Nair MS, Revathy K, Tatavarti R (2008) Removal of salt-and-pepper noise in images: a new decision-based algorithm. In: Proceeding of the International MultiConference of Engineers and Computer Scientists, vol 1, IAENG, Hong Kong, pp 19–21

9. Ng P-E, Ma KK (2006) A switching median filter with boundary discriminative noise detection for extremely corrupted images. *IEEE Trans Image Process* 15:1506–1516
10. Wang Z, Bovik A, Sheikh H, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 13:600–612

# Chapter 42

## Research and Design of Process Data Warehouse for Business Process Assessment

Hui Xia, Qing Yao and Fei Gao

**Abstract** Business process optimization occupies a very important position in any corporation, the root of optimization lies in the mining of process logs. However, the log data in different information systems is often heterogeneous, for ease of process mining, they must be processed together in an united way, and data warehouse (DW) technology is the best choice. Appropriate process warehouse structure becomes the key to the research. Based on existing business process assessment model, this paper designs a process assessment-oriented process warehouse for storage and management of process instance data. It can be applied to the process mining, assessment and optimization of event logs. Finally, a case study demonstrates the concept given by this paper.

**Keywords** Data warehouse · Process log · Process assessment · Process Warehouse

### 42.1 Introduction

Business process optimization has become the core competitiveness of enterprise operation, and information system is the main pillar of business process. They, such as ERP, CRM, SCM and so on, have achieved the systematic management for

---

H. Xia (✉) · Q. Yao (✉)  
Shan Dong University, Jinan, Shandong, China  
e-mail: xiahui1005@gmail.com

Q. Yao  
e-mail: yaoqing@sdu.edu.cn

F. Gao  
Equipment Department of 71146 Army, Weifang, Shandong, China  
e-mail: 185641016@qq.com

the business process in the corporation. When these information systems are running, they record all the operation information into the system event logs. With the passage of time, a great quantity of information related to process execution exists in the logs, and it's an important resource for business process mining and optimization. The problem that different information system may has heterogeneous or redundancy information is not conducive to the business process mining, assessment and optimization. Data warehouse can integrate the information in each data source, and makes it the basis of data analysis. A data warehouse built for the analysis of business process is called Process Warehouse [1]. With the help of OLAP tools, it can implement information aggregating, analysis, and comparing, moreover, it can mining new process model and improve the quality of the existing process model.

Building a process warehouse will face lots of challenges: analyst may have different abstract level and data granularity; synchronization between process analysis and process automation; different information system has diverse life cycle, furthermore status number in life cycle is infinite [2]; the relationship between dimension tables and fact tables; inhomogeneity of items in fact tables; the interchangeability of dimension tables and fact tables; diversity and so forth.

Based on the existing process assessment models, this paper proposes a structure of process data warehouse which is more generic, process assessment-oriented, and optimization-oriented.

## 42.2 Related Work

According to Inmon's definition in the literature [3], Data Warehouse is a subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management's decisions. Paper [4] breaks through the traditional design method of Data Warehouse, proposes an unified representation of model, and problems comes down to that of the choice of a hierarchy of criteria adapted to the necessities of analysis. It also elaborates the model method of fact table and dimension table. In order to solve the problems encountered in the process analysis [2], proposes a generic solution for process warehousing and a process warehouse model. On the basis of [2], paper [5] solves ETL (Extract-Transform-Load) problems and dumps log data into process warehouse.

In [6], it introduces a generic data warehouse design for processing workflow log data, there are a lot of workflow logs in workflow management systems (WFMSs), and they are a very valuable resource for business process re-engineering and optimizing. Business process-oriented data warehouse architecture is proposed in [7–9], and [7] deduces the structure of the data warehouse through the business process model. The capabilities of process warehouse are seldom evaluated, so [10] develops a generic process analysis framework (PAF) that can be used for evaluating analysis capabilities of a process warehouse. Based on both theoretical and technological achievements in processing mining fields, [11] puts forward a process



mining model suitable for event log mode, provides the analysis process of selecting evaluation indicators by AHP, and gives out a set of formulas for assessing mining instance combing with the vector model.

On the basis of existing business process assessment model, this paper takes process assessment-oriented process warehouse technologies as a study object. Through preprocessing process logs, we can load them into process warehouse, then it will be convenient for extracting useful process information for the optimization of the business processes basing on the assessment method [11].

## 42.3 Process Warehouse

In this paper, the Data Warehouse built for business process analysis is called Process Warehouse (PW). It is the major object in following analysis, which provides data source for process mining.

### 42.3.1 Process Log

In information systems, Process logs have detailed record of the execution of activities in process instance. It has a vital role as the data source of process warehouse. The process log in this paper adopts XES, which is based on XML, thus eliminating the need for evolution in process assessment model. The event here refers to the action defined in process model, and instance is the once execution of process.

At present, the main object of process mining is structured event log presented by symbols. However, a great quantity unstructured logs exist in practical applications, so a very important work is the structuralization and normalization of logs before the data loaded into process warehouse [10]. Although the structure of log in different information systems is different, the basic information they record is similar, Table 42.1 is a typical structure of process log.

**Table 42.1** A fragment of process log

Instance id	Event id	Timestamp	Activity	Executor	Cost
1	13110041	2012.12.109:02	Register request	Mike	10
	13110042	2012.12.109:09	Examine	John	50
	13110043	2012.12.114:56	Check	Pete	100
	13110044	2012.12.208:32	Decide	Sara	100
2	13110045	2012.12.112:21	Register request	Ellen	10
	13110046	2.12.12.113:01	Reinitiate request	Mike	20

### 42.3.2 Process Assessment Model

In order to do process mining better in business intelligence environment, appropriate process mining models and assessment method for process mining results become key issues. Reference [11] proposes a process mining model used for event log mode, and it gives out a set of formulas for assessing process instance. Results of assessment are presented as numerical data, can be displayed directly for its significance. Through the setting of threshold, we can effectively select and classify the process instance, thus providing the basis for decision making for the follow-up process optimization.

Figure 42.1 shows the process assessment model, which uses the XML language to record process event based on the form of event log and evolves referencing XES standards. The relationship of elements in the model is shown following.

Process log is the source of process mining, assessment and optimization. The PAM based on the mining of logs provides an important reference for the

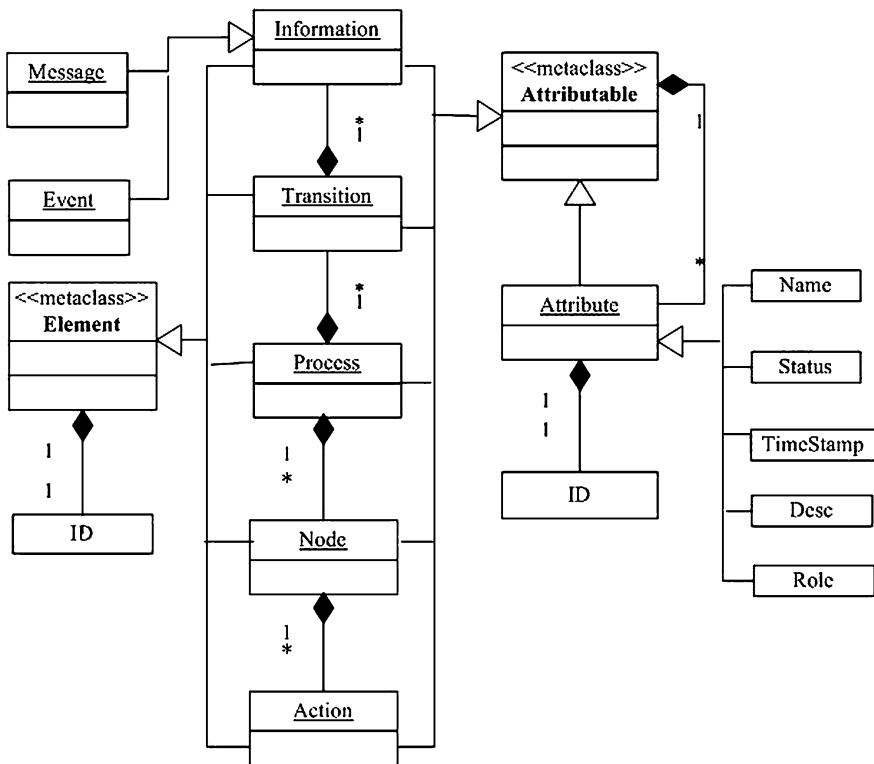


Fig. 42.1 The process assessment model

subsequent optimization. As the storage component, process warehouse plays a very good role of bridge between the log and assessment model.

### 42.3.3 Design of Process Warehouse

Process warehouse based on the PAM stems from data warehouse, and here proposes a generic and process assessment-oriented process warehouse model (AOPWM). A single fact table is adopted in this paper, and its theme is the execution of process. To make it as granular as possible is suitable for a variety of data analysis.

Figure 42.2 shows the main elements of the snowflake schema of multi-dimensional AOPWM, and the formalization describes as follows:

**Definition 1** AOPWM is a fact-dimension mode, specified by a five-tuple  $(Fs, Ds, R, MD, BD)$ .  $Fs$  is the set of facts, and  $Ds$  is the set of dimensions;

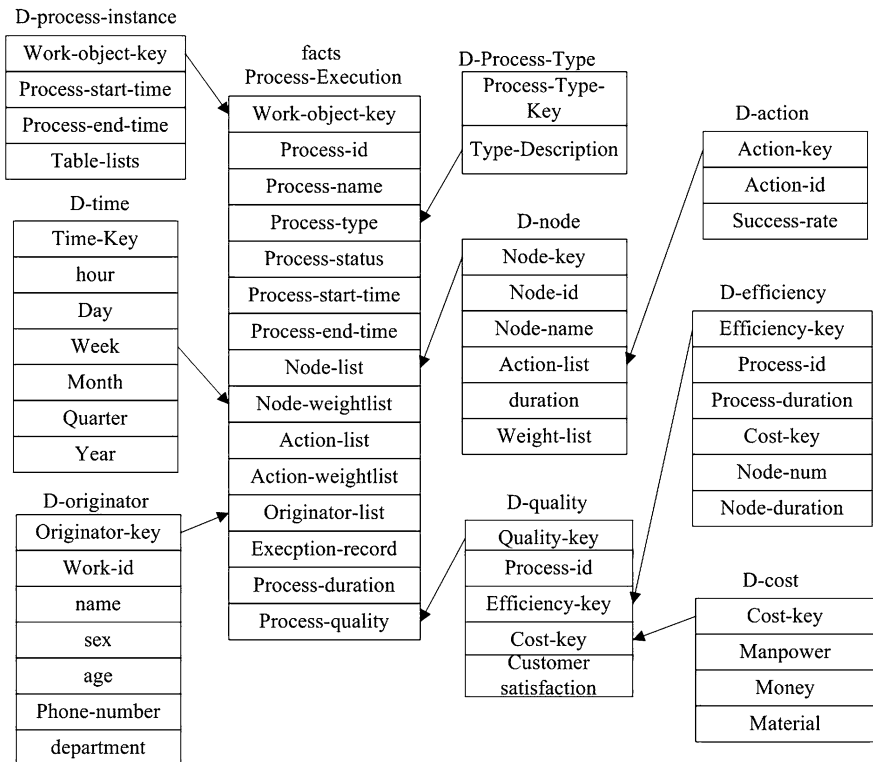


Fig. 42.2 Access-oriented process warehouse model

$R$  represents the relationship between facts and dimensions;  $MD$  is metadata of warehouse, and  $BD$  is the related business data.

In this paper, we use a matrix to express the relationship between facts and dimensions as well as dimensions and dimensions. We assume that AOPWM has only one fact and  $n$  dimensions, then a matrix can be used to represent it. Here the last row and column is fact, and others represent dimensions. For example,  $r(n+1)j = 1, j \in (0, 1, \dots, n)$  means the fact referencing  $j$  dimension, and if the number is 0, then the reference relation does not exist. The paper only considers the reference which is from fact to dimension, and rule that dimensions can only reference dimensions, and does not appear loop.

**Definition 2** A fact is a five-tuple  $(F, key, PA, RA, Measurements)$ , where:  $F$  represents this is a fact;  $key$  is a two-tuple  $(id, name)$ ;  $PA$  is the set of primary attributes;  $RA$  is a set of attributes which reference to a member of dimensions;  $Measurements$  is a set of numeric attributes, which includes various values.  $RA$  corresponds to the last row of  $R$  in AOPWM, which specifies the relation between fact and dimensions. If here use  $\alpha n + 1$  as last row of  $R$ , then  $\sum \alpha(n+1) = ||RA||$ .

**Definition 3** Dimension, a four-tuple  $(D, key, PA, [RA], [Property])$ , where:  $D$  represents this is a dimension; The means of  $key, PA$  and  $RA$  are similar to the ones in fact;  $Property$  shows the features of this object,  $[]$  means this attribute is optional.

In process instances, it can establish different tables according to different process types. The *quality* dimension makes it easy for evaluators to find the properties to assess, such as *efficiency, customer satisfaction, cost* and others. Here the *quality* dimension references attributes in *efficiency* dimension and *cost* dimension, and according to the importance of *efficiency, customer satisfaction* and *cost*, it can set a weight for them, calculates a numeric result of quality, which could intuitively represent the quality.

The *efficiency* of the process execution is defined as the number of nodes that are executed within unit time. *Customer satisfaction* is the percentage that the number of satisfied nodes accounts for the total in this process, and it is a very crucial performance indicator for enterprises, which can directly reflect the result of process to some extent. The *cost* dimension contains human, financial, material and other aspects of attributes, makes evaluators easily find the needed data. The *time* dimension has been throughout the entire fabric, plays a vital role, and can do a variety of statistical analysis using aggregate functions.

The process loaded into process warehouse during ETL is completed, here does not consider the uncompleted instance. Process belonging to different types may be executed more than once, and each execution will add one record in facts and the instance table at the same time. Due to the different data sources, it may lead to different integrity of each attribute in instance, in addition, the attributes and status of each instance are numerous, but in practical application, the properties enterprise concerned is limited, which makes our model feasible.

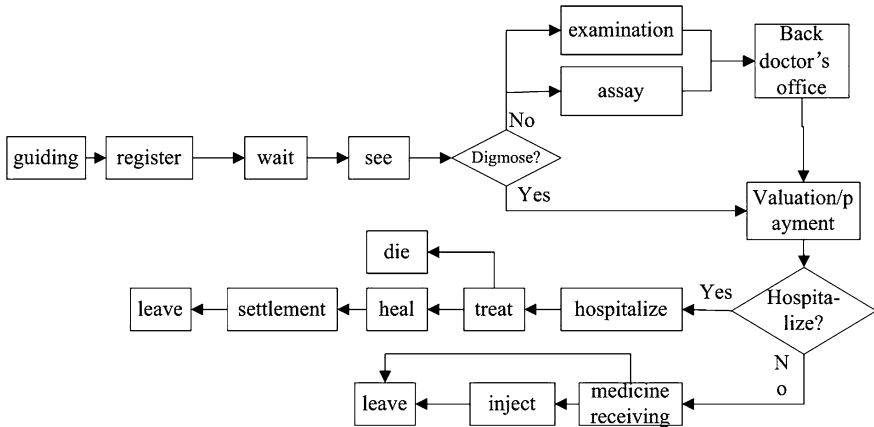


Fig. 42.3 A flow chart of medical circuit for breast cancer

### 42.4 Case Study

Here uses a case to demonstrate the concepts defined in this paper. Taking the data of some cancer hospital for example, it has a total of 374 patients with breast cancer: in the hospital lobby there exists guiding place; then patients register, wait and see the doctor. Doctors will write checklist according to patients' condition, then valuating and doing the corresponding examination. Figure 42.3 is the flow chart. For shortcutting, one node hosts only one action.

After executing several healthcare processes, they will leave a large number of records about patients' information in the log, like name, age, sick type. Table 42.2 is a fragment of information in system.

The fact (Table 42.3) will increase one record when instance 1 is loaded into the process warehouse. Here Table 42.4 is the quality dimension.

Here makes the following statistic analysis for the instances in data set: the maximum of efficiency is 1.71 and the minimum is 0.48; for cost, they are 58649 and 12652; for patient-satisfaction, they are 1.0 and 0.48.

Because of the difference between value scale of attributes in Table 42.4, it can not be used to calculate quality directly, so normalization is needed at first, here using *min-max normalization*:

$$y = \frac{x - MinValue}{MaxValue - MinValue} (New\_MaxValue - New\_MinValue) + New\_MinValue \tag{42.1}$$

*MaxValue* and *MinValue* are the maximum and minimum; *x* is the value before standardization and the result is *y*, mapping *x* to new interval [*New\_MinValue*, *New\_MaxValue*].

**Table 42.2** Information for patients

Case-id	Name	Age	Type	Tumor size (mm)	Node caps	Inv nodes	Deg malig	Irradiat	Breast
1	Hong Zhang	67	Benign	34	No	0	1	Io	Right
374	Ling Zhao	52	Benign	30	No	0	1	No	Left

**Table 42.3** Fact table

Case-id	Type	Status	Start time	End time	Node list	Action list	Origin ator	Duration	Quality
1	Benign	Completed	2012/09/18	2012/09/28	Guiding ..... Leave	Guiding ..... Leave	Hong Zhang	10 days	6.272

**Table 42.4** Dimension table of quality

Instance-id	Efficiency	Cost	Patient-satisfaction
1	1.2	30000	0.83

Here uses formula 42.1 to normalize *efficiency* and *satisfaction* respectively, since the value of cost is the smaller the better, we do following transformation for the formula 42.1:

$$y = \frac{MaxValue - x}{MaxValue - NewValue} (New\_MaxValue - New\_MinValue) + New\_MinValue \tag{42.2}$$

For instance 1 in Table 42.2, mapping the values of efficiency attribute and satisfaction to [0, 10] and we get the following result:  $y_{efficiency} = 5.85$ ,  $y_{satisfaction} = 6.73$  and the normalized result of cost is  $y_{cost} = 6.23$ . Then according to  $quality = \sum y_i w_i$ , ( $i = 1, 2, 3$ ), it can calculate *quality* based on the *weights* set by business personnel. Here the value of  $i$  corresponds to the three attributes in quality dimension, and  $w_i$  is the weight of them, and here they are 0.35, 0.3, 0.35. The result of quality is 6.272 after calculating.

## 42.5 Conclusions

In this paper we propose a design method of process assessment-oriented process warehouse, give out the content of process logs and the structure of process warehouse. In process warehouse, the determination of facts and dimension tables as well as the attribute in them are all process assessment-oriented, which is conducive to the following mining and assessing work. At last, using several medical circuit instances, we explain how to store them and how they can be used for process assessment by using method proposed in this paper. Now the process warehouse structure is not overall, and it needs further optimization.

## References

1. Matthias J, Thomas L et al (2000) The challenges of process data warehousing. In: VLDB '00 Proceedings of the 26th international conference on very large data bases, pp 473–483
2. Casati F, Castellanos M, Dayal U, Salazar N (2007) A generic solution for warehousing business process data. VLDB'0, Vienna, Austria, pp 23–28
3. Inmon WH (2003) Building the data warehouse, 3rd edn. Wiley, New York
4. Schneider M (2007) A general model for the design of data warehouses. *Int J Prod Econ* 112(2008):309–325
5. Johann E, Georg EO, Wolfgang G (2002) A data warehouse for workflow logs. In: EDCIS '02 Proceedings of the first international conference on engineering and deployment of cooperative information systems, pp 1–15
6. Böhnlein M, Ende AU (2000) Business process oriented development of data warehouse structures. *Data warehousing 2000, Methoden, Anwendungen, Strategien*; Physica: Heidelberg, 2000
7. Doria D, Feldmana R, Sturmb A (2008) From conceptual models to schemata: an object-process-based data warehouse construction method. *Inf Syst* 33(6):567–593
8. Sturm A (2012) Supporting business process analysis via data warehousing. *Journal of Software: Evolution and Process, Special Issue: Business Process Modeling, Development and Support*, 2012, 24(3):303–319
9. Shahzad K, Johannesson P (2009) An evaluation of process warehousing approaches for business process analysis. In: EOMAS '09 Proceedings of the international workshop on enterprises and organizational modeling and simulation, article no. 10
10. Geng L, Buffett S et al (2009) Discovering structured event logs from unstructured audit trails for workflow mining. *Found Intell Syst* 5722:442–452
11. Wang Z (2012) The research of process mining assessment used in business intelligence. *computer and information science (ICIS)* In: Proceedings of the 2012 IEEE/ACIS 11th international conference on, pp 179–183

## Chapter 43

# A Novel Emergency Cross-Media Information Retrieval Model

Lingling Zi, Junping Du, Qian Wang and Jangmyung Lee

**Abstract** Existing information retrieval approaches provide only limited capabilities to capture the query requirements. However, a complete understanding of search requirements is essential for improving the effectiveness of retrieval in the emergency management field. To achieve this goal, we proposed a novel emergency cross-media information retrieval model, which includes four parts: information collection, information indexing, information retrieval and intelligent mobile terminal. The proposed model has two advantages. One is to use ontology technique to identify appropriate semantic information according to query words. The other is to use image semantic analysis based on SIFT to achieve the task of emergency image retrieval. Conducted experiments show that our model obtained encouraging performance results.

**Keywords** Emergency information · Cross-media · Information retrieval · Ontology · Query expansion

---

L. Zi · J. Du (✉) · Q. Wang  
Beijing Key Lab of Intelligent Telecommunication Software and Multimedia,  
School of Computer Science, Beijing University of Posts and Telecommunications,  
Beijing 100876, China  
e-mail: junpingdu@126.com

L. Zi  
School of Electronic and Information Engineering, Liaoning Technical University,  
Huludao 125105, China

J. Lee  
Department of Electronics Engineering, Pusan National University, Busan 609735,  
Korea



## 43.1 Introduction

With the increasing of multimedia information, requirements for emergency information retrieval are growing [1, 2]. At the present, traditional retrieval modes are confronted with big challenges: (1) a lot of uncorrelated information is easy to return; (2) the key word-based retrieval mode provides limited capabilities to capture user implicit query need; (3) the accurate multimedia search results cannot be found for users.

In face of this situation, we proposed a novel emergency cross-media information retrieval model (EIRM). The main contributions of the proposed model are as follows: (1) ontology technique is adopted to analyze user implication requirements and more appropriate query expansion words can be obtained; (2) the technique of image semantic analysis based on SIFT is introduced to the field of information retrieval, and emergency semantics can be extracted reasonably; (3) a pattern that enables unified multimedia resources (i.e. text resources and image resources) in the emergency domain is developed. In a word, EIRM integrates the advantages of keywords and semantic search and supports more accurate emergency multimedia search results, thereby improving the performance of retrieval.

The rest of the paper is structured as follows. [Section 43.2](#) presents the framework of EIRM. [Section 43.3](#) illustrates the implementation of EIRM. [Section 43.4](#) presents experimental work to demonstrate our model. The last section concludes the paper.

## 43.2 The Framework of EIRM

In this section, we demonstrate a cross-media information retrieval model for emergency information. This model consists of four components (See Fig. 43.1).

- (1) Information collection module: it is responsible for collecting the massive emergency information from the Internet, including text collection, image collection, text extraction and semantic annotation. The objects of collection are the text and image information in the emergency management. We define collection keywords depending on the type of emergencies, and use a collection algorithm based on content evaluation, which can determine the degree of association between collected information and emergencies. Meanwhile, the texts are extracted from collected text and image information and then semantics are annotated.
- (2) Information indexing module: aiming to quickly search emergency multimedia information, we need to build up the index in the model. Specifically, the function of this module contains three parts: the annotated document parsing, inverted index creation and index maintenance. The first part performs the

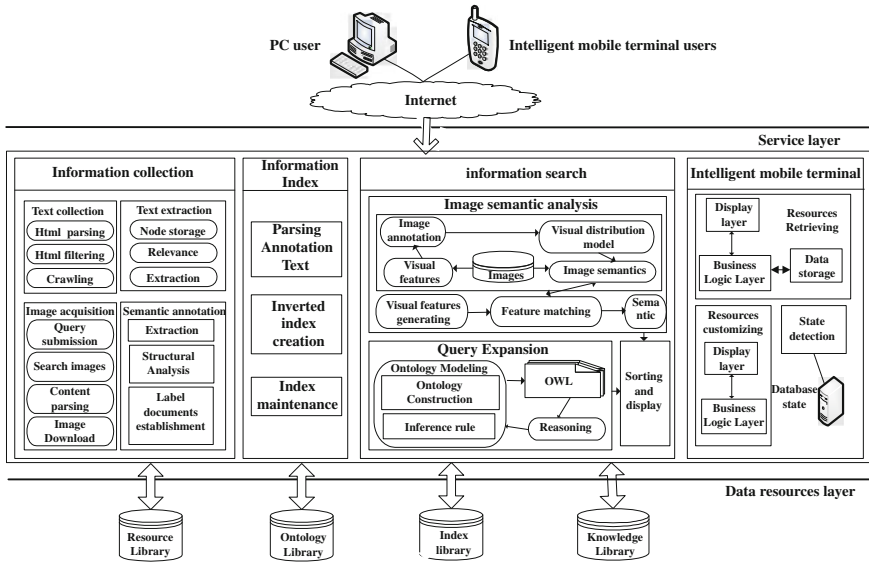


Fig. 43.1 The framework of the proposed model

parsing task for annotated documents, the second part creates inverted index for the parsed documents and the last part updates the index for changed annotated document by batch updating and incremental updating.

- (3) Information retrieval module: this module is responsible for accurate search emergency cross-media information, which mainly contains three parts: image semantic analysis, query extension based on ontology, and sorting and display. The first one performs the acquisition of image semantics by extracting the underlying features of the query image; the second one achieves query extension using emergency ontology and knowledge reasoning; the last one proposes the label sorting method to obtain comprehensive emergency information query results, combining text and image resource from multiple angles.
- (4) Intelligent mobile terminal module: it is responsible for searching and displaying emergency information in mobile terminal. It is divided into three parts: resource customizing, resource retrieval and state detection. The first one selects the collected resources, maintains data object lists for mobile terminal retrieving, and sets the display list of retrieving results for different user groups. The second one calls information retrieval module in the servers, carries out heterogeneous information retrieval, and displays results based on the customized information provided from back-end database. The last one adopts resource locating function to monitor regularly database objects, and records monitoring results in the log files.

### 43.3 The Implementation of EIRM

In this section, we elaborate the critical steps in the process of EIRM implementation. The proposed model is not only able to capture accurately the user query intention, but also to provide multi-faceted emergency multimedia search results.

#### 43.3.1 Inverted Index Construction

EIRM adopts inverted index technique [3] and information caching technique to construct index files, so as to improve the efficiency of retrieval system. Specifically we firstly analyze the established label documents, and extract the index terms, including metadata index term and content index term. Then the index fields are established according to the index terms, and the inverted index files can be constructed. Here the inverted index file is denoted as  $\langle k, \langle a_1, f_1, \langle p_{11}, p_{12}, \dots, p_{1f_1} \rangle \rangle, \dots, \langle a_i, f_i, \langle p_{i1}, p_{i2}, \dots, p_{if_i} \rangle \rangle, \dots, \langle a_k, f_k, \langle p_{k1}, p_{k2}, \dots, p_{kf_k} \rangle \rangle \rangle$ . In which  $k$  represents the number of annotating words appearing in the label documents,  $a_i$  is the ID of label document. Given the label document  $a_i$ ,  $f_i$  is the term frequency of query word and  $\langle p_{i1}, p_{i2}, \dots, p_{if_i} \rangle$  is its position list.

The inverted file list can be divided into three parts: (1) a ID sequence of label document  $\langle a_1, \dots, a_i, \dots, a_n \rangle$  ( $1 \leq i \leq n$ ); (2) a position sequence  $\langle p_{i1}, p_{i2}, \dots, p_{if_i} \rangle$ ; (3) a term frequency sequence  $\langle f_1, \dots, f_i, \dots, f_k \rangle$ . Given the query word, all label documents associated with the query word, i.e.  $\langle a_1, \dots, a_i, \dots, a_n \rangle$  are quickly found through inverted index files and the corresponding files can be obtained using the obtained label documents. Finally the source files of the various types of media information can be acquired according to the information recorded in the label documents.

#### 43.3.2 Image Semantic Analysis Based on SIFT

In order to ensure the accuracy of image analysis, EIRM combines feature mapping and text extraction to complete image query function of emergency cross-media information. The semantic analysis processes are shown as follows.

**Step 1:** the candidate image set can be obtained using image collecting function and visually similar images in the set can be acquired through applying visual characteristics method.

**Step 2:** the key words can be extracted from the sources of the acquired images and from obtained key words [4]; the global key words are captured as the texts of image annotation. Accordingly, all the used images are in the training set of semantic annotation.

**Step 3:** image SIFT features [5] can be computed and a visual distribution model can be constructed in these training images. On this basis, a mapping between the text annotation and visual distribution [6] can be established by applying machine learning method.

**Step 4:** K-means clustering algorithm [7] is used to cluster the feature vectors of all training images. Each cluster is treated as a visual block and each training image eigenvectors are distributed in the various clusters. Finally, image semantic description is obtained according to the obtained distribution of visual blocks.

When the user enters an image, EIRM generates visual features of the image, and performs matching of visual characteristics based on the training sets. Finally according to matching results, we obtain semantic description of the input image.

### 43.3.3 Query Extension Based on Ontology

To realize query expansion, EIRM uses ontology technique [8, 9] to represent emergency knowledge, performs knowledge reasoning through emergency ontology reasoning, and effectively shares and reuses emergency knowledge. We use object-oriented approach to design conceptual model of emergency knowledge and the model contains three elements: concept, attribute and relation. Concept refers to the name of the basic unit of emergency ontology; attribute refers to the characteristics of emergency knowledge, and relation describes the association between emergency knowledge.

Specifically, emergency ontology can be established according to the emergency conceptual model, including establishment of concept, attribute and relation. Then inference rules are designed based on the principles of OWL and the inference engine can be created according to description information and ontology. Meantime, inference engine and query ontology are bound together to obtain a search object. Combing the ontology API and model API, search object can be reasoned and so new knowledge can be acquired. When the user inputs query keywords, semantic expansion words are acquired through new knowledge and more searching results are displayed for uses.

### 43.3.4 Label Sorting

We use the label sorting method to organize the searching results according to the correlation of the query expansion set and the annotation information. The specific processes are shown as follows:

**Step 1:** calculate the correlation between expansion words and result records. Let  $E = \{e_1, e_2, \dots, e_n\}$  be the extended word set. The degree of correlation between expansion word  $e_i$  and the label document, i.e.  $Rank(e_i, label)$  is computed according to (43.1):

$$\text{Rank}(e_i, \text{label}) = \begin{cases} \sum_{j=1}^{\text{Occurance}(e_i, \text{label})} \ln \frac{\text{Length}(\text{label})}{\text{Location}(e_i, j, \text{label})}, \text{Occurance}(e_i, \text{label}) > 0 \\ 0, \text{Occurance}(e_i, \text{label}) = 0 \end{cases} \quad (43.1)$$

where  $\text{Length}(\text{label})$  represents the length of the label document;  $\text{Occurance}(e_i, \text{label})$  represents the frequency of  $e_i$  that occurs in the label document;  $\text{Location}(e_i, j, \text{label})$  represents the location that  $e_i$  occurs in the label document. Then the correlation between extended word set and the label document, i.e.  $\text{labelrank}(E, \text{label})$  is computed using (43.2).

$$\text{labelrank}(E, \text{label}) = \sum_{i=1}^n \text{Rank}(e_i, \text{label}) \quad (43.2)$$

**Step 2:** determine expansion degree of  $e_i$ , i.e.  $\zeta$  according to the position of its result list of ontology reasoning.

**Step 3:** calculate the final correlation between  $E$  and label documents by using (43.3):

$$R(E, \text{label}) = \text{labelrank}(E, \text{label}) \times \zeta \quad (43.3)$$

Finally, multi-faceted emergency information search results integrated with text and image can be sorted by  $R(E, \text{label})$  and shown for users in the navigation view.

## 43.4 Experiments Results and Discussion

We constructed emergency cross-media information retrieval system for users who query in Chinese inborn language. For the development of this system, we used Myeclipse 8.5 platform, MySQL 5.1 and a PC with Intel Core(TM) 2 Duo T6570 processor, 2.1 GHz and 4 GB of main memory. Lucene and Heritrix web crawler tool were also used. In addition, we built emergency ontology with Protégé tool and performed ontology reasoning with Jena.

Precision and Recall are important measurements of the search system. We used ECIR to perform five queries, including milk poisoning (query ID 1), Red Duck (query ID 2), Watermelon leavening (query ID 3), waste oil (query ID 4) and shoes gelatin (query ID 5). The precision and recall values corresponding to the results of different queries are shown in Fig. 43.2. It can be seen that all the precision values are 100 % and recall values are 80 %, 70 %, 90 %, 78 % and 89 % respectively. The results demonstrate that the performance of EIRM is relatively stable.

We investigated the system performance from the perspective of the user with correct results provided by humans. For this reason, ten students from our department were asked to use this system. The volunteers entered the specified query keywords. Ranking accuracy and satisfaction score were recorded according to the results returned. Figure 43.3a depicts the average ranking accurate rate of

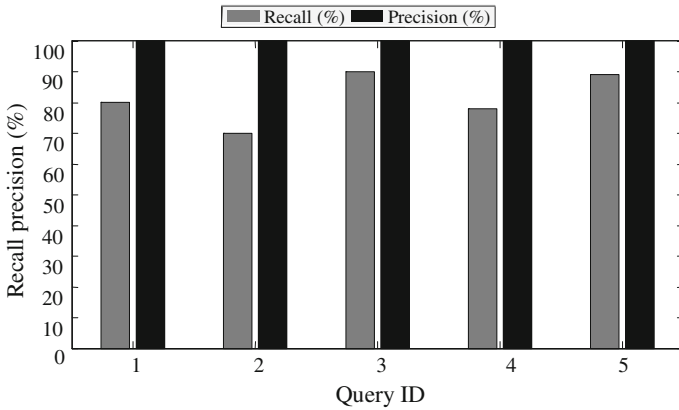


Fig. 43.2 Precision and recall results

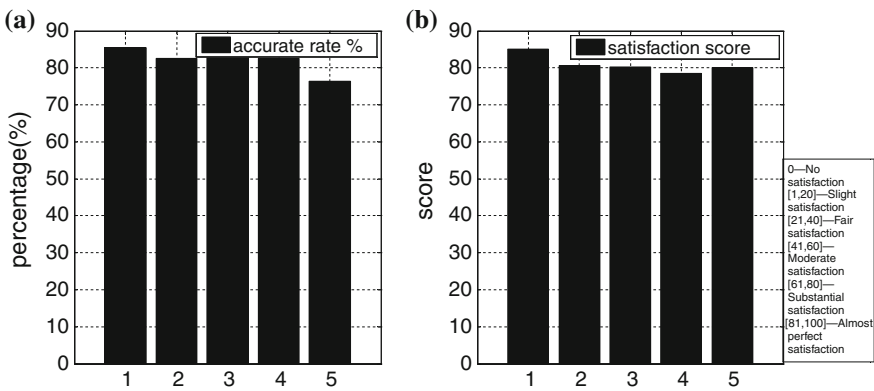


Fig. 43.3 Performance evaluation by users

our survey and Fig. 43.3b summarizes the volunteers' average satisfaction score with regard to query results. The satisfaction standards of grading are shown on the right. The average satisfaction score is 81 and it demonstrates that users are relatively satisfied with the query results. However, in our survey, there are also some cases of relatively low satisfaction score, possibly due to because the fact that some multimedia objects are not marked accurately.

### 43.5 Conclusions

This paper proposes a novel emergency cross-media information retrieval model. The proposed model collects emergency multimedia information, performs cross-media semantic analysis and achieves efficiency retrieval for cross-media

information. The model also achieves the task of user query under the state of emergency management by using the retrieval of intelligent mobile terminal, which adapts to unstable network condition in the emergency areas. Further work includes intelligent query of emergency cross-media information, emergency forecasting and intelligent decision.

**Acknowledgments** This work is supported by the National Basic Research Program of China (973 Program) 2012CB821200 (2012CB821206), the National Natural Science Foundation of China (No. 91024001, No. 61070142) and the Beijing Natural Science Foundation (No. 4111002).

## References

1. Carpineto C, Romano G (2012) A survey of automatic query expansion in information retrieval. *ACM Comput Surv* 44(1):1:1–1:50
2. Ferrandez A (2011) Lexical and syntactic knowledge for information retrieval. *Inf Process Manage* 47(5):692–705
3. Kucukyilmaz T, Turk A, Aykanat C (2012) A parallel framework for in-memory construction of term-partitioned inverted indexes. *Comput J* 55(11):1317–1330
4. Kallipolitis L, Karpis V, Karali I (2012) Semantic search in the world news domain using automatically extracted metadata files. *Knowl-Based Syst* 27:38–50
5. Dorado-Munoz LP, Velez-Reyes M, Mukherjee A, Roysam B (2012) A vector SIFT detector for Interest point detection in hyperspectral imagery. *IEEE Trans Geosci Remote Sens* 50(11):4521–4533
6. Zhou GT, Ting KM, Liu F (2012) Relevance feature mapping for content-based multimedia information retrieval. *Pattern Recogn* 45(4):1707–1720
7. Lee SS, Lin JC (2012) An accelerated K-means clustering algorithm using selection and erasure rules. *J Zhejiang Univ-Sci C-Comput Electr* 13(10):761–768
8. Valencia-Garcia R, Garcia-Sanchez F, Castellanos-Nieves D, Fernandez-Breis JT (2011) OWLPath: an OWL ontology-guided query editor. *IEEE Trans Syst Man Cybern Part A Syst Hum* 41(1):121–136
9. Fernandez M, Cantador I, Lopez V (2011) Semantically enhanced information retrieval: an ontology-based approach. *J Web Seman* 49(4):434–452

# Chapter 44

## A Framework for 3D Model Acquisition from Multi-View Images

Chunmei Duan

**Abstract** 3D model acquisition is a fundamental issue in computer graphics and computer vision. However, constructing 3D model manually using software such as 3D MAX and Maya is a tedious and expensive work. Therefore, finding out how to obtain 3D model directly from the real world becomes a hot research topic. In this paper, we describe a framework for obtaining 3D model from multi-view images of a real object. We start with images of an object taken from different views, and then feature points extracted and matched. From the correspondences, camera calibration data and 3D geometry are acquired. Experimental results in the end of the paper show the effectiveness of the framework.

**Keywords** 3D reconstruction · Feature correspondence · Surface reconstruction · 3D structure recovery

### 44.1 Introduction

3D model acquisition is an important task in industry, and also is a fundamental issue in research fields of computer graphics and computer vision. However, to construct 3D model manually using software such as 3D MAX and Maya is a tedious and expensive work. Therefore, finding out how to obtain 3D model directly from the real world becomes a hot topic in research fields. Currently, as a digital reservation and record technology, 3D structure acquisition from real objects can be widely applied to fields of object modeling, scene modeling, photorealistic rendering, robot navigation, object recognition and 3D metrology and other cultural fields such as archaeology, advertising and entertainment.

---

C. Duan (✉)

School of Management Science and Engineering, Shandong Normal University,  
East Wenhua Road, Jinan 250014 Shandong, People's Republic of China  
e-mail: chunmeiduan@mail.sdu.edu.cn



Methods for 3D model acquisition from real objects can be mainly categorized as active and passive approaches. Active methods are used by 3D scanner community and passive method usually defined by 3D modeling based on images. Due to its low cost and immediately color acquisition power, image-based 3D modeling becomes complement of active methods.

Automatically reconstructing 3D structure from multiple images taken from different views is also known as image-based modeling and is a key research topic in the automated acquisition of geometric object models. In this paper, aiming at acquisition of integrated 3D model, we exploit the whole workflow of 3D reconstruction and present a simple and low-cost framework for 3D modeling of real world object from multi-view images.

## 44.2 Related Work

Over the past years, a great many of related studies aiming at constructing 3D models from images existed [1–9]. Approaches for obtaining 3D data from images can be generally classified as four categories. ① Voxel-based approaches [3, 4] are based on a bounding box containing the reconstructed scene and their accuracy is limited by the resolution of the bounding box. ② Methods based on deformable polygonal meshes [5, 6] need starting with a good point such as a visual hull model, which limits their applicability seriously. Esteban and Schmitt [6] propose a method using visual hull to initialize the deformation of a surface mesh under the influence of photo consistency constraints, which yields excellent results except for some local minima. ③ Approaches employing multiple depth maps [7] are more flexible alternatives, however, its procedure need complicated depth map fusion. ④ Small patches or points based methods [1, 8] retrieve 3D structure from images by extracting small components such as points or patches which construct the final 3D model in combination. Furukawa and Ponce [1] propose a novel algorithm based on a dense set of small patches covering the surfaces visible in the multi-view images and the experimental results are impressive.

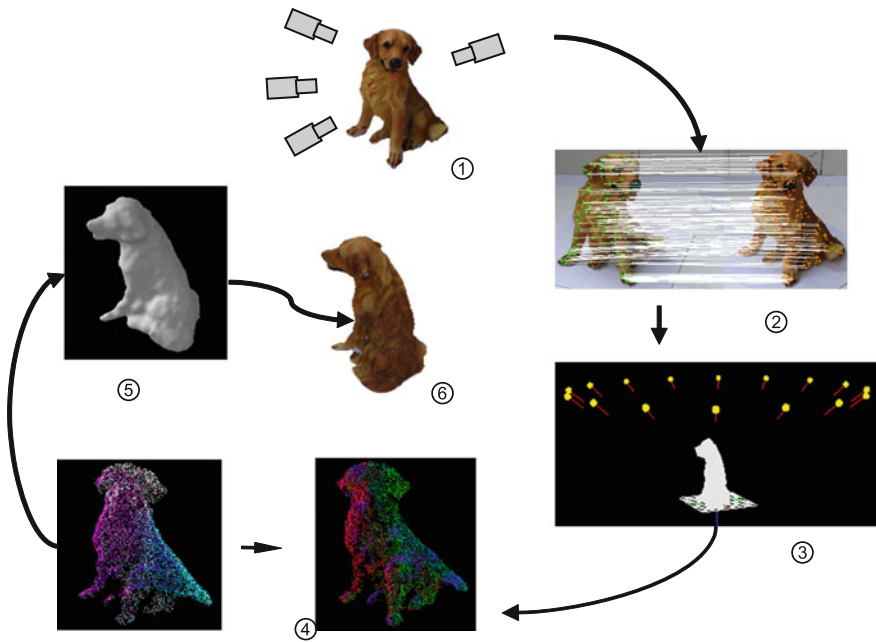
Although 3D model acquisition from multi-view images is a well studied topic in past years, it still suffers from some problems such as demand for expensive experimental equipment for obtaining high quality image, ideal illumination conditions and accurate camera calibration data, etc. In this paper, we present a simple, low-cost and hybrid framework for 3D model acquisition from images. We first start with a few images of a real world object taken from different views by a single camera, and then feature points of images extracted and matched. The matches are used for camera calibration which can be employed to produce a dense set of feature matches. After calculation, 3D point cloud is generated from the collection of image matches. Final 3D models are obtained after 3D point cloud optimization, surface reconstruction and multi-view texture-mapping.

### 44.3 The Framework for 3D Model Acquisition from Multi-view Images

Figure 44.1 shows our system for 3D model acquisition of real world object from multi-view images. The framework start from step ① images acquisition, step ② local image feature extraction and matching, step ③, ④ 3D point set retrieving and optimization, step ⑤, ⑥ surface reconstruction and texture mapping. The rest of the section describe these steps concretely.

#### 44.3.1 Acquisition of Multi-view Images

Firstly, as shown in Fig. 44.1 step ①, we take photos of a real world object from different views. As a matter of fact, in order to avoid the vibrations of the camera, we just attach it to a sturdy tripod and turn the target object every time for a new view. The first two columns of Fig. 44.3 show two views out of total 16 images taken from all around of three test objects. The photos are acquired by digital camera Canon A640 and the image resolution is  $640 \times 480$ . The entire photo taking procedure is implemented by computer software provided by Canon Company.



**Fig. 44.1** Our framework for 3D model acquisition of real world objects from multi-view images

### 44.3.2 Image Local Feature Extraction and Matching

Then, as shown in Fig. 44.1 step ②, images feature points of every views should be extracted and matched for estimation of camera intrinsics and extrinsic parameters used for 3D structure recovery. We detect the scale invariant features known as Difference-of-Gaussian (DoG) [10] which is defined as the difference of Gaussian convolution kernel of different scale factors as described in Eq. (44.1).

$$D(x, y, \sigma) = (G(x, y, k\sigma)) - G(x, y, \sigma) * I(x, y) \quad (44.1)$$

where  $G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$  is the Gaussian kernel,  $I(x, y)$  is the processed image and  $*$  denotes the convolution operation.

Then the gradient data of the neighborhood patches in suitable size of the DoG features are normalized and then the local feature descriptor described in [11] is applied to perform feature correspondence task. The feature correspondences are used to perform camera motion parameter estimation.

### 44.3.3 3D Point Set Retrieving and Optimization

The matched features are firstly used for estimating fundamental matrix [9] which describes the relationship between different image views and is estimated non-linearly using global optimization and bundle adjustment techniques [12] jointly. The fundamental matrix is parameterized to the minimum number of seven parameters and estimated by global minimization in term of non-convex linear matrix inequality (LMI) and convex LMI relaxation techniques [13]. We perform the computation in a RANSAC framework and consider nonlinear criteria, together with epipolar geometry constraint [9].

The multi-view projective reconstruction in a unified framework is only based on the fundamental matrix, by which the 3D structure is created. Based on the solved fundamental matrix, the projection matrices in projective space corresponding to different views are estimated. The solved projective matrices and 3D point set in projective space can be updated to the metric space by self-calibration technique [9]. Due to the robustness and preciseness of the estimated fundamental matrices, our method to recover the projection matrices and 3D structure is stable and robust. The procedure is referred to as the step ③ in Fig. 44.1.

After imposing the dense matching similar to Ref. [1], we obtain relatively dense 3D points. However, due to the ill-posed nature of the algorithm, the obtained 3D point cloud is sparse and noisy comparing with those obtained by 3D scanners. In 3D reconstruction field, conventionally, bundle adjustment (BA) will be adopted to optimize 3D point clouds and camera calibration data [9]. During the minimization shown in Eq. 44.2, the 3D point structure and parameters of camera projection matrix are optimized. There are enough correspondences in adjacent two and three views according to statistics we got in experiments and we divide all

the image views into 2-view and 3-view groups and over which to perform optimization algorithm.

$$\min_{\{P_i\}, \{M_j\}} \sum_{j=1}^N \sum_{i=0}^T \{m_{ij} - f(P_i, M_j)\} \quad (44.2)$$

where,  $\{P_i\}$  is the collection of projection matrices,  $\{M_j\}$  is the collection of the estimated 3D points,  $\{m_{ij}\}$  is the collection of matched feature points of  $M_j$  under point view  $i$ ,  $\| \|$  is 2-norm operator, namely, sum of squares,  $f$  is the projection function, namely,  $f(P_i, M_j) = P_i M_j$ .

We implement the optimization using SBA technique. SBA (Sparse Bundle Adjustment) is an effective development of LM algorithm which employs the sparse nature of coefficient matrix so as to simplify the solving [14]. The refined correspondences and the estimated projection matrices as well as the initial recovered 3D point set are as the parameters being sent to SBA framework to perform local and global optimization. RANSAC framework is also used to increase the robustness of the algorithm. This part is referred to as the step ④ in Fig. 44.1. Our experimental results show that errors in the original data are corrected and the quality of the reconstructed 3D model is promoted obviously.

#### 44.3.4 Surface Reconstruction and Texture Mapping

After optimization, 3D point clouds will be triangulated to be a 3D mesh model. In our experiments, we first employed Tight Cocone method [15] which is based on Delaunay triangulation. Although Tight Cocone is very effective to scanned data, it

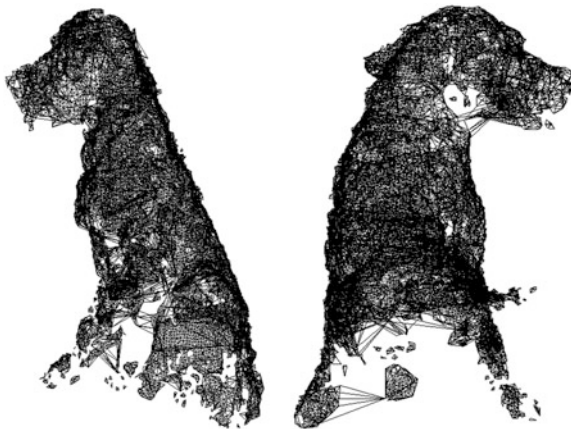


Fig. 44.2 3D mesh model after Tight Cocone triangulation

**Table 44.1** Part of data sets used in our experiments

Data sets	Resolution	Illumination	Image number	Materiel
Dog	640 × 480	Natural light	16	Resin
Warrior	640 × 480	Natural light	16	Bronze
Beauty	640 × 480	Natural light	16	Wooden

**Fig. 44.3** Three reconstructed 3D models obtained from multi-view images

is not very good at dealing with sparse and noisy data like ours. As shown in Fig. 44.2, many errors and holes are found on surface of the mesh after Cocone triangulation. So, we adopt Poisson approach [16] to triangle the oriented points. And the final 3D points are further smoothed after employing the approach as shown in Fig. 44.1 step ⑤.

Texturing the reconstructed 3D model is the last step in the framework as shown in Fig. 44.1 step ⑥. We propose a new method for automatically texturing complex 3D model in [17]. The photographic images for acquiring 3D structure are again used as texture data and mapped precisely on the surface to enhance the

model appearance. We implement an iteration algorithm for automatically sampling a few images as texture maps instead of using the whole set of input images.

## 44.4 Experimental Results

Table 44.1 shows the characteristics of part of the data sets that we used in our experiments. Figure 44.3 shows three reconstructed 3D models obtained from multi-view images using the proposed framework. The left two columns show two of the 16 images of real objects for reconstruction. The rest columns show the recovered point sets and textured 3D models, respectively.

## 44.5 Conclusion

In this paper we presented a framework for automatically construct 3D model of real world objects from multi-view images of the objects. The integrated 3D reconstruction workflow was revisited and in each step of the reconstruction procedure such as feature matching, point cloud generation, optimization and texture mapping. Some new approaches or algorithms were employed to promote the final quality of the reconstructed 3D model. We believe that the proposed framework provided an effective and inexpensive way to obtain 3D model of real world objects. The experimental results in the end of the paper show the efficiency and feasibility of the proposed algorithm.

**Acknowledgment** This research was supported by the National Nature Science Foundation of China (61170038), a Project of Shandong Province Higher Educational Science and Technology Program (J13LN14) and the Open Project Program of the Shandong Provincial Key Lab of Software Engineering (2011SE003), China.

## References

1. Furukawa Y, Ponce J (2010) Accurate, dense, and robust multi-view stereopsis. *IEEE Trans Pattern Anal Mach Intell* 32:1362–1376
2. Schwing AG, Hazan T, Pollefeys M, Urtasun R (2012) Efficient structured prediction for 3D indoor scene understanding. In: *IEEE conference on computer vision and pattern recognition*
3. Pons JP, Keriven R, Faugeras OD (2007) Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *Int J Comput Vision* 72(2):179–193
4. Sinha S, Mordohai P, Pollefeys M, Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh. In: *IEEE international conference on computer vision*
5. Furukawa Y, Ponce J (2008) Carved visual hulls for image-based modeling. *Int J Comput Vision* 81(1):53–67

6. Hern'andez Esteban C, Schmitt F (2004) Silhouette and stereo fusion for 3D object modeling CVIU 96(3):367–392
7. Bradley D, Boubekeur T, Heidrich W (2008) Accurate multi-view reconstruction using robust binocular stereo and surface meshing. In: IEEE conference on computer vision and pattern recognition
8. Lhuillier M, Quan L (2005) A quasi-dense approach to surface reconstruction from uncalibrated images. IEEE Trans Pattern Anal Mach Intell 27(3):418–433
9. Hartley RI, Zisserman A (2004) Multiple view geometry in computer vision, 2nd edn. Cambridge University Press, Cambridge
10. Lowe D (1999) Object recognition from local scale-invariant features. In: IEEE International conference on computer vision
11. Duan C, Meng X, Tu C (2008) How to make local image features more efficient and distinctive. IET Comput Vision 2(3):178–189
12. Triggs B, McLauchlan P, Hartley R et al (1999) Bundle adjustment—a modern synthesis. Lecture Notes in Computer Science. pp 298–372
13. Kahl F, Hartley R (2008) Multiple view geometry under the L (Infinity)-norm. IEEE Trans Pattern Anal Mach Intell 30(9):1603–1617
14. Lourakis M, Argyros A (2009) SBA: a software package for generic sparse bundle adjustment. ACM Trans Math Softw 36(1):1–30
15. Dey T, Goswami S (2003) Tight Cocone: a water-tight surface reconstructor. J Comput Inf Sci Eng 3(4):302–307
16. Kazhdan M, Bolitho M, Hoppe H (2006) Poisson surface reconstruction. In: Eurographics symposium on geometry processing
17. Duan C, Meng X, Tu C (2010) Optimization of multi-view texture mapping for reconstructed 3d model. In: Proceedings of the 8th world congress on intelligent control and automation, pp 30–34

# Chapter 45

## Vehicle Tracking Based on Nonlinear Motion Model

Fan Zhang, Hong Li, Kalilou Kone and Wei Zhang

**Abstract** How to detect and track vehicle in nonlinear motion is a big problem in computer graphics technology. Monocular vision is complete works like positioning with only one camera; it has a simple structure and a strong applicability. In this paper, we build and solve models of traffic video with nonlinear motion model, to locate and track vehicles in video. We present a new nonlinear movement model using shape model and projection model. The accurately calculated results of position and velocity of the vehicle is given by estimating the vehicle's speed, posture, the last position. It's an effective solution to the problem that nonlinear movement is difficult to estimate by KALMAN filter and the tracking process is not stable and reliable enough. The experiments show that the new algorithm has better robustness than EKF and UKF algorithm.

**Keywords** Nonlinear motion · Vehicle tracking · Motion model · KALMAN filter

### 45.1 Introduction

With the development of science and the progress of the society, the car has become an indispensable transport tools. Vehicles detection and tracking in the traffic video is the basic step of intelligent transportation system, visual monitoring and it is also very important step. In recent years, both attach great importance to

---

F. Zhang (✉) · H. Li · K. Kone  
School of Information Science and Engineering, Central South University, Room 314,  
Comprehensive Experiment Building, Railway Campus, Changsha 410012, China  
e-mail: zfjz123@csu.edu.cn

W. Zhang  
School of Architecture, Hunan University, Changsha 410082, China



the research of this field at home and abroad, and successively appeared a lot of remarkable achievements. For most of the tracking problem, it usually assumes that the camera is stationary, or compensate through a global movement. The key to the problem is foreground and background modelling, and the model must meet the following requirements:

1. Model of the background and the foreground has sufficient separability.
2. Between adjacent frames, the measured value of the parameter in the foreground model has good continuity and stability.
3. The model is able to meet the requirements of real-time computing; otherwise, the motion prediction model will not be able to maximize its effectiveness, leading to the failure of the tracking.

Mean-shift, Particle filter and KALMAN filter tracking method has become widely used in recent years, but the description of the foreground/background has been key to the study [1, 2]. The central idea is in the case of priori geometric model of the vehicle, according to the information of its position, size, and speed in a single view, determined the next time location, size, angle, and other parameters of the estimator, efforts to find the optimal solution of the foreground object, In order to achieve a stable and powerful tracking result. In addition, there are two non-linear motion tracking method: The first method is the function approximation.

It is for linearization of the nonlinear equation of state or observation equation, typical algorithm is extended KALMAN filter (EKF) [3]. The EKF First linearizes nonlinear equations, using a Taylor expansion and then combines the classic KALMAN filter for the estimation. EKF algorithm is simple and easy to implement, but has poor tracking performance in the strongly nonlinear and non-Gaussian environment, and even filtering divergence. Another method is based on the sampling approximation method, sampling approximation method using a set of samples with weighting value to approximate the posterior probability density function (PDF) of the target, typical algorithms are Unscented KALMAN filter (UKF) and the particle filter (PF). Which UKF [4] use the identified sample points, Thus avoiding tracking errors caused by the linearization, when the state of the posterior probability density function is non-Gaussian, the tracking performance will get worse. The experiment references EKF and UKF methods.

## **45.2 Motion Model**

### ***45.2.1 Linear Motion Model***

Firstly, this paper describe the most simple model of rigid linear motion like that the motion of the vehicle only in a straight line, only have one degree of freedom. On this basis, discuss it further more in depth and complex. In the situation which

the motion of the vehicle is linear entirely, we suppose that the target vehicle's projected image in monocular video is  $Y$ , and its profile is described by a series of coordinate points, the parameter of the profile is  $S = \{x_1, y_1, x_2, y_2, \dots, x_N, y_N\}$ ,  $x_i, y_i$  are the coordinate points which were marked as group  $i$ .

If we know the profile  $S_\tau$  at  $\tau$  time, we can get the description value at  $\tau + 1$  time:

$$S_{(\tau+1)} = S_\tau \times s_\tau + t_\tau \quad (45.1)$$

The expression of  $t_\tau$  like that

$$t_\tau = v_\tau \times \Delta\tau + 1/2 \times a_\tau \times (\Delta\tau)^2 \quad (45.2)$$

$v_\tau$  and  $a_\tau$  Originate from the estimated value in  $\tau$  time,  $t_\tau$  present the instantaneous displacement of the object in  $\tau$  time,  $s_\tau$  presents the scale changes which were caused by displacement. In the linear motion, the motion of the object do not cause any change in its shape and gesture, so the scale changes is proportional to the visual depth of its location, and the visual depth is also linear when the motion is linear. So we can calculate  $s_\tau$  from the gradient.

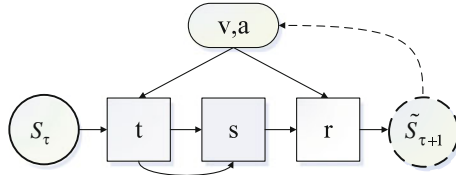
### 45.2.2 Nonlinear Motion Model

As the nonlinear motion of the rigid body, we can view it as the motion in two directions—translation and rotation. Translation like the linear motion, just expand the motion in straight line to the whole plane or carved surface, it cause the change in coordinate and scale. Rotation is the projection which the moving object in three dimensional space projects its rotation on two-dimensional plane, it mainly cause the tracking object's change in shape and gesture. We also should know that, in the nonlinear motion of the vehicle, the generation of the rotation is accompanied by its displacement; there also have some certain relationships between the angle of rotation and the parameters of the displacement. Now we are discussing the nonlinear motion model in plane.

$$S_{(\tau+1)} = r_\tau(S_\tau) \times s_\tau + t_\tau \quad (45.3)$$

$r_\tau$  is the rotation which caused by moving object, and it is also a process which caused by the two-dimensional profile's changes in shape and gesture. In the flat surface,  $t_\tau$  of nonlinear is same as linear model, but  $s_\tau$  is defined by the visual depth of the object's coordinate location. In general, vehicle always towards the tangent direction of its movement, so, in the flat surface, the angle of the  $r_\tau$ 's rotation is  $(a_\tau \Delta\tau / 2v_\tau)$ . Besides that, because of the vehicle always on the land, the horizontal angle of the vehicle is also the horizontal angle of land in the picture.

Then we analysis these parameters further more. Firstly it's the translations parameter  $t$ , as the most important attribution in vehicle tracking. We can describe



**Fig. 45.1** The generative motion model. The graph implies three types of information: the translation model  $P(t|v, a)$ , the scaling model  $P(s|S, t)$  and the Rotation model  $P(r|S, v, a)$ .  $S$  will get the estimated value with these three models, and give the feedback to the parameter  $v$  and  $a$

it with  $v$  and  $a$ .  $E(v) = v_0 + a_0\Delta$ , we can consider the speed to be linear, in a short time  $\Delta \rightarrow 0$ . But  $v_0$  and  $a_0$  originate from the estimation of the previous information, so,  $v$  also has some certain Gaussian noise.

$$a \propto N\left(\frac{dv}{dt}, \sigma^2\right) \tag{45.4}$$

If we don't have priori knowledge, we can consider a have additive Gaussian noise. Scale scaling parameter  $s$  for Plane, the scaling coefficients in the x-axis and y-axis  $S_x, S_y$  is completely independent,  $s = s_x \times s_y$ . For the curved or irregular terrain,  $s = s(\tau + 1)/s_\tau$ ,  $s(\tau + 1)$  and  $s_\tau$  are determined by the coordinate and the topographic gradient. The rotation parameter  $r$  rotates the 3D object, start with the angle  $\theta_\tau$  and end with the angle  $\theta(\tau + 1)$ . The motion model like Fig. 45.1.

## 45.3 Vehicle Tracking with Nonlinear-Model KALMAN Filter

### 45.3.1 Initialization

Before the tracking, we should detect the vehicle. When an object enters the camera screen, detection starts immediately. The principal framework of vehicle detector is composed of 3 parts: firstly, we extract edge features of the input image, especially extract the profile features of the new object; then combine the profile with the original picture and cut the object apart; finally, send the segmentations results to the image classification device, then we can get the detection results. If the new object is vehicle, the tracking aim, extract its angle features; discern the profile and the direction of the vehicle. The finally results of the detector have 3 parts: vehicle profile, the rectangle framework of the vehicle location, vehicle direction; send these information to the tracking device, then the tracking start.

### 45.3.2 The Tracking Method

We apply the motion model which proposed previously to the target tracking, select KALMAN filter as the general framework of the tracking, form a complete tracking module. In this experiment, define the status of KALMAN filter as  $x(t) = [x, y, dx, dy, d^2x, d^2y]^T$ , the observation is  $y(t) = [x, y]^T$ ,  $x/dx$  and  $d^2x$  are target image's location in horizontal direction, velocity and acceleration;  $y/dy$  and  $d^2y$  are target image's location in vertical direction, velocity and acceleration. The basic tracking procedure is as follows:

1. Initialization phase: define the observation target  $z_p$  in the reference image and input its initial status  $x(0)$ . Select the process model of KALMAN filter and forecast the region where the target possibly appears.

$$x(t) = A \cdot x(t-1) + w(t) \quad (45.5)$$

which

$$A = \begin{bmatrix} 1 & 0 & \Delta t & 0 & \frac{1}{2}\Delta t^2 & 0 \\ 0 & 1 & 0 & \Delta t & 0 & \frac{1}{2}\Delta t^2 \\ 0 & 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 1 & 0 & \Delta t \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\Delta t = 1, p(w) \propto N(0, \sigma^2 I), \sigma^2 I = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Forecast the location of the object with the formula, so we get the estimated information  $x(t)$  at  $t$  time. If the accuracy of the forecast model is not enough, do once template matching after we get the estimated value of  $x(1)$ , acquire the accurate information about velocity and location.

2. Renew the estimation of the target's location: according to the detection method, acquire the target's observation location  $z_p(t)$ , renew the estimation of the target's status.

$$x(t) = \tilde{x}(t) + K(t)(y(t) - H \cdot \tilde{x}(t)) \quad (45.6)$$

In the formula,  $H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$ ,  $K(t)$  is KALMAN gain.

- Correct the status of the tracking target with the template in motion model, correct the tracking framework and optimize the tracking barycenter with the object's location estimation  $x(t)$  and angle estimation  $\arctan \left[ \frac{(d^2x + d^2y)\Delta t}{2(dx + dy)} \right]$ , then we can get the estimation  $x(t + 1)$ .

## 45.4 Experimental Results

### 45.4.1 Accuracy Analysis

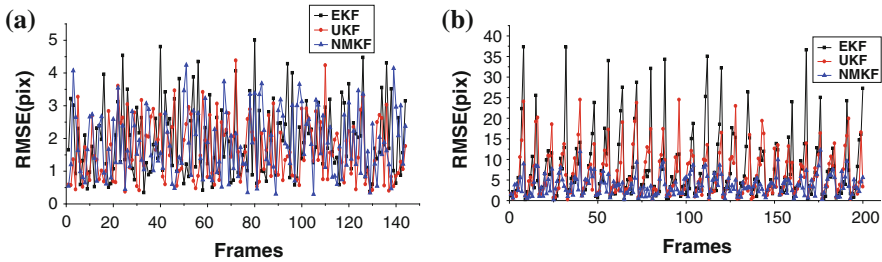
We compare our approach with EKF and UKF. We initialize the filter with inputting the first two frames' annotation of each moving vehicles.

Figure 45.2 shows examples of tracking with three methods EKF, UKF and NMKF. It is a typical example of a turning movement. Since the turn, the vehicle's shape, speed, and other parameters are changed. EKF still trying to find best estimate then caused large deviations. UKF try to produce smooth results with posterior probability using Gaussian distribution. However, the error is still beyond the Gaussian noise model accommodated upper limit. So UKF results are not accurate enough. The new method does more accurate tracking by matching motion model. The result of new approach only has a tiny distance with the observed value.

We focus on the non-linear movement, which contains 200 frames. For each example, we run EKF, UKF, and our approach separately, using the same initialization and background difference method as detection. Figure 45.3 shows the errors for the three methods. The x-axis shows index of frames. The y-axis shows the root mean square error (RMSE) with respect to the labelled point.



**Fig. 45.2** Nonlinear object tracking result from different tracking method. The EKF algorithms use black box, The UKF algorithms use red box. Our NMKF algorithms use yellow box. **a** EKF. **b** UKF. **c** NMKF



**Fig. 45.3** Test error for vehicle tracking with EKF, UKF and our NMKF approach. **a** Linear movement. **b** Nonlinear movement

**Table 45.1** Time complexity of nonlinear motion tracking

Filter	The average running time (ms)	RMSE (pix)
EKF	15.2	10.6
UKF	90.6	7.5
MMKF	166.0	3.3

### 45.4.2 Speed Analysis

We research the tracking of non-linear motion; the average time this experiment takes (Each trace of the mean value of each tracing time) is shown in the Table 45.1.

As can be seen by the table information, the computational complexity of NMKF increases. However, adaptability for nonlinear movement enhances. It can still effectively track in complex circumstances, it is more robust than other algorithms.

## 45.5 Conclusion

This paper proposed a new motion modelling method of the core issues in the target tracking, Modelling and description of the target, and improved the traceability of the nonlinear target motion and the stability of the target tracking. The experiment results showed that: This approach can be effective tracking nonlinear motion object in the long sequence of frames; this tracking prediction model is relatively independent with shape modelling method. Although this paper uses KALMAN filter as tracking framework, other tracking methods are possible. It has good extensibility. All of these experiments show that the modelling method proposed in this paper can enhance the performance of the tracking algorithm.

## References

1. Li L, Huang W, Gu IY, Qi T (2004) Statistical modeling of complex backgrounds for foreground object detection. *IEEE Trans Image Process* 13(11):1459–1472
2. Wang H, Suter D (2007) A consensus-based method for tracking: Modelling background scenario and foreground appearance. *Pattern Recognit* 40(3):1091–1105
3. Julier SJ, Uhlmann JK (2004) Unscented filtering and nonlinear estimation. *Proc IEEE* 92(3):401–422
4. Arulampalam MS, Maskell S, Gordon N (2002) A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Trans Signal Process* 50(2):174–188

# Chapter 46

## Detecting Pedestrian Using Motion Information and Part Detectors

Lingli Xu and Zhiping Zhou

**Abstract** A pedestrian detection method based on motion information and part detectors is proposed in this paper used for handling partial occlusions of pedestrian in video. Extracting motion areas in the video image by fast frame difference image Gaussian mixture model as the candidate region of the pedestrian firstly; Then the part detectors including head, head-left should, head-right should, torso and so on, which have trained by the liner SVM combined with the HOG features pyramid were used to scanning detect in each candidate region individually. Finally, the Max Margin Hough Transform is used to verify the detection result. Experiments on databases and the video shoot by us show that our method has high performance in detecting pedestrians with partial occlusion.

**Keywords** Partial occlusion · Motion information · Part detectors · Pedestrian detection · HOG features pyramid

### 46.1 Introduction

Pedestrian detection has very important applications in video surveillance, machine vision, assisted driving, and content-based image/video retrieval. However, detecting humans in images/videos is a challenging task because of the variation in visual appearance caused by changes in pose, clothing, articulation and occlusion. Currently, the most classical approach is the Histograms of Oriented Gradients (HOG) which proposed by Dalal et al. [1] in 2005. A great test result can be got by this method when the background of image is relatively simple and the target does not be occluded. However, there often exist various occlusions between background

---

L. Xu (✉) · Z. Zhou

School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China  
e-mail: xllzya@sina.com



and pedestrians, pedestrians and pedestrians in actual scenes. It is widely accepted that a representation based on parts is necessary to tackle the challenge of detecting people in images. Mykhaylo and Stefan [2] built upon the pictorial structures model and strong part detectors for people detection and pose estimation, which is powerful and general. But it did not take full advantage of model relationships between body parts beyond kinematics constraints. Felzenszwalb et al. [3] generalized an approach for object detection based on mixture of multi-scale deformable part models combined with latent SVM, which was a better performing system on the task of pedestrian detection. However, the part models did not consider the effect of detection affected by difference of angles of parts of human body. Guo et al. [4] described a segmentation-aware model to handle occlusion in object detection, which made the detection more robust to occlusion, but it need to richer representation on the deformable parts-based models.

## 46.2 Getting the Part Detectors

### 46.2.1 *Selecting Parts of Human*

The low-level signals captured by HOG features pyramid from the image blocks are often ambiguous owing to the variation of the posture and angle of human and the irregular occlusion; patterns that are almost indistinguishable given the HOG features pyramid inside the image blocks, but can be distinguish by observing the addition shape context of human [5]. For example, a front-facing pattern is similar to a back-facing one in appearance, and thus the corresponding part detectors will often have the similar test results. If the separate parts of the body is used to split human that human body is can be divided into head, arms, torso, legs simply. Using these separate parts for training to detect pedestrians will bring a relatively large error. These ambiguities can be resolved by exploiting context. To distinguish between the pedestrian together through multiple patterns of neighboring parts detectors. For example if a head detector detects only, we can infer that the oval shape such as a wheel sometimes will be mistaken for a head in the image. But, if use a head-and-shoulder pattern, we can suppress the false detection by determining whether there is no torso underneath.

As we as know, there are 19 types of key-point in anatomical human (such as eyes, nose, hip, joints, etc.), and 15 types of regions of person [6] can be obtained, which include face, left should, right leg and so on. For pedestrians walking upright, the approach of key-points annotating from different angles of head, left and right should, hip is used to fragmenting the 30 individual body parts for training in this paper, such as head, left head-should, front-torso, upper body, etc. The image patches including these parts regarded as positive samples combined with negative samples that are the same size image blocks randomly generated from non-pedestrian samples can be trained to produce 30 types of parts detectors.

### 46.2.2 Training Local Detectors

The difficulty and the primary problem is how to divide the parts and construct effective part detectors based on local detection. Parikh et al. [7] found that the detection performance impacted on parts detectors is greater than the impacted of the geometric relationship between the parts detectors on local detection. Therefore, the most importance is constructing effective parts detectors. There the dataset we used is the PASCAL VOC 2011 which provides enough such annotations. In the dataset the human body is treated as a lot of image blocks corresponding to different parts of person. The training algorithm consists of the following steps:

**First step:** Collecting image blocks. 30 random windows contain the key-points of the human body is selected from the training set as seed windows. Patches are extracted from other training examples with similarly local key-points configuration with each seed windows. Then following the Eqs. (46.1) and (46.2) to compute a distance between the set of  $K_1$  composed of the all key-point of seed window and the key-point configuration  $K_2$  within the annotated image of people. The image block will be discarded whose distance is too large. The distance metric is proposed following:

$$D(K_1, K_2) = D_{proc}(K_1, K_2) + \lambda D_{vis}(K_1, K_2) \quad (46.1)$$

$$D_{proc}(K_1, K_2) = \sum_{i=1}^n \left[ (x_{i1} - x_{i2})^2 + (y_{i1} - y_{i2})^2 \right] \quad (46.2)$$

where  $D_{proc}$  is the Procrustes distance between the two common sets and  $D_{vis}$  is a visibility distance. The two configurations can be ensured have a similar aspect effectively by  $D_{vis}$ .

**Second step:** Training and classifier. The collected patches are regarded as positive samples. The HOG features pyramid is extracted from the positive samples and random negative samples, and 30 parts classifiers can be trained by linear SVM classifier. To ensure that all training blocks are adequately close to the seed window, only the nearest 300 training examples are used in this paper. According to standard practice, initially training SVMs by scanning over images excluding people, retain by collecting hard false positives.

The following figure shows a part of the collected training image blocks and trained the local detectors, respectively corresponding to the different parts of the human body (Fig. 46.1).

For walking person, the head and torso of persons are relatively stable compared with other parts. So many literatures only use a single head [8] or torso for training to detect pedestrians. The performances of these methods are effective when targets do not be occlusion, but the detection performance will decline significantly when there occlusion exists between object. Figure 46.2 shows the effect of only using head or torso detector detecting pedestrians.



**Fig. 46.1** The parts of the human body and its corresponding partial detector. The *top row* shows the image patches of various parts of body collected within the dataset, as positive samples for training. The corresponding to parts detectors are shown in the following row. They are respective front-face, front-face-double-shoulder, front-face-right-torso, side-torso, front-torso, and front-half-face

As shown in the figure, the pedestrians with occlusion cannot be accurately detected by single detector. If the results cannot be received by some detectors when the human body is partially occluded, can we use others to obtain the results. Therefore the problem of partial occlusion can be solved by combining a plurality of detectors without affecting the final detection results, which can significantly improve the detection rate.



**Fig. 46.2** Part of the experimental effect diagram. **a, b** Shows respectively detecting no-occluded pedestrians with a single head and torso detector, which can accurately detect people; and **c, d** shows respectively detecting occluded pedestrians with the same detector, which will cause undetected

### 46.3 Extracting Candidate Regions

If we use all parts detectors had trained before to detect scanning over the whole image waiting to detect the time consuming of the pedestrian detection task may be difficult to accept. Thence the two-steps contain pedestrian candidate region assumed and objectives confirmed are used in this paper. Pedestrian in the name suggest is a walking person, so the motion areas from the images can be captured as the candidate regions of the pedestrian. For complex outdoor scenes, targets can be detected effectively by the Gaussian Mixture Model. A novel method named FDGMM to solve this problem by modeling the frame differencing image with fast update GMM.

Let to:  $(x, y)$  is the cording of a pixel in  $I(x, y, t)$  which is the  $t$ -th sequence of video image. And  $Y(x, y, t)$  is the gray value of  $I(x, y, t)$ . Then inter-frame differencing image is introduced to replace the gray image to built new Gaussian Mixture Models. The differencing image is generated as:

$$d(x, y, t) = |Y(x, y, t) - Y(x, y, t - 1)| \quad (46.3)$$

And DGMM can be shown this:

$$P(d_t | d_{t-1}) = \sum_{i=1}^K \omega_i * \eta \left( d_t \middle| u_{i,t}, \sum_{i,t} \right) \quad (46.4)$$

where  $K$  is the number of Gaussian models, which is empirically to be 2 in this paper,  $\omega_i$  are the weights which satisfy  $0 < \omega_i < 1$  and  $\sum \omega_i = 1$ , and  $\eta(X_t, \mu_{i,t}, \sum_{i,t})$  is the normal probability density of mean  $\mu_{i,t}$  and covariance  $\sum_{i,t} = \sigma_{i,t}^2 I$ .

According to the illustration, the parameters of Gaussian mixture models must be initialized by calculating the mean and covariance of inter-frame differencing image sequences. But the more frames we used the greater amount of computation will be taken. There we selected the first 20 frames initializing the algorithm.

Then he GMM should be updated with background pixels real-time used for reducing the influence of illumination conditions and the false positive alarm. Unlike the traditional GMM algorithm, in this paper, two independent and continuously updatave learning rates are used to respectively update mean and variance quickly [9]. The learning rates for mean calculated as follows:

$$\alpha_k(t) = \alpha_k(t - 1) + \frac{K + 1}{K} \eta_k - \frac{1}{K} \sum_{i=1}^K \eta_i \quad (46.5)$$

The initial value is  $\alpha_k(0) = 0.05$ . As can be seen from the formula, the learning rate is increased if the probability of pixel belonging to the  $k$ th Gaussian component is higher than average probability of others. Then the update of the mean is this:

$$\mu_k(t) = (1 - \alpha_k(t))\mu_k(t-1) + \alpha_k(t)X \quad (46.6)$$

Here a function  $\beta_k(X, \mu_k)$  is introduced for the variance to ensure a quasi-linear adaptation. The function is following:

$$\beta_k(X, \mu_k) = a + \frac{b-a}{1 + e^{-s\varepsilon(X, \mu_k)}} \quad (46.7)$$

where  $\varepsilon(X, \mu_k) = (X - \mu_k)^T(X - \mu_k)$  the parameter  $s$  is equal to 0.005, and  $a, b$  limit the variance value to a domain  $D \in (\frac{a+b}{2}, b)$ . Then the variance is updated independently from the mean as this:

$$\sigma_k^2(t) = (1 - \xi)\sigma_k^2(t-1) + \xi \cdot \beta_k(t) \quad (46.8)$$

where the  $\xi = 0.6$  is a constant learning rate.

## 46.4 Locating and Segmentation in Hypotheses

The candidate region is needed to further confirm whether it is pedestrian and determine the location of the pedestrian. In this method, a Generalized Hough transform voting structure is used to achieve the positioning of pedestrians.

Each part detectors run at every location and scale of the input image to collect all possible for the key-points of the human body, which is called activation point. For example the  $i$ -th activation of the  $p$ -th local detector can be represented as  $p_i$  and the score of classifier assigning is  $a_i$ . Selecting image blocks with a fixed size around these activation points, as the windows may contain a pedestrian portion. Then we use mean shift algorithm to cluster all windows. The target position will be casted a vote by Each cluster The probability of detecting at  $x$  of object  $O$  is calculated as:

$$P(O|x) \propto \sum_i \omega_i a_i(x) \quad (46.9)$$

where,  $\omega_i$  is the weight of the part detectors. The peaks of Hough space can be find by compute the cast votes and the sum over the detectors according to Eq. (46.9) in each cluster. Then the Max Margin Hough Transform (M<sup>2</sup>HT) [10] that can find the set of weights which minimize the false positive but maximize the true positive is used to learn the weights. We use the  $\omega_i = 1$  in formula to calculate the peaks in Hough space in the training set. The optimization mechanism of the weights is following as this:

$$\min_{w, b, \xi} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^T \xi_i \quad (46.10)$$

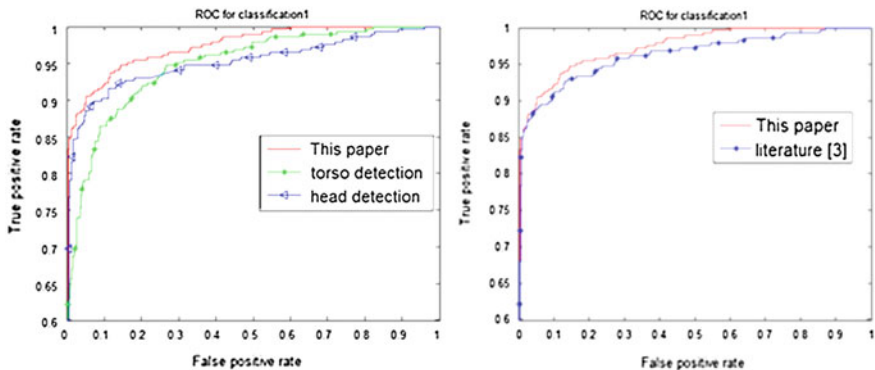
$$\begin{aligned} S.t. & y_i(\omega^T A_i + b) \geq 1 - \xi_i, \\ & \omega \geq 0, \xi_i \geq 0, \forall i = 1, 2, \dots, N \end{aligned} \quad (46.11)$$

where  $A_i^j$  is the score of  $j$ -th local detector within the  $i$ -th Hough peak. If  $y_i = 1$  the peak is true positive, and  $y_i = -1$  if false positive. This can accurately detect the local position of people.

## 46.5 Experiments and Analysis

The pedestrian detection problem is mainly discussed when the target exists obscured. Thence, a method combined with a lot of parts detectors to deal with the problem of target occlusion. If pedestrians can not be detected by some detectors we can use others. To further illustrate the problem, the method of this paper, a single head detector, a single torso detector is used respectively to test the test set in INRIA dataset which contain 288 positive samples and 402 negative samples drawing the ROC curve, as show as Fig. 46.3. The literature [3] also used the method combined with many local detectors to detect pedestrians. But it neither considered the difference of angles of parts of human body, nor consideration of combining with multiple parts of a person. However, the parts detector are obtained by selecting different angles of one or a plurality of continuous parts of human for training in this paper. Similarly, the two methods are tested in the test drawing a ROC curve, as show as Fig. 46.4.

Obviously, the left ROC curve of Fig. 46.3 demonstrates that the detection results of the proposed method in this paper are superior to the other two methods. From right of Fig. 46.4 can also be apparent to the detection results of this method are better than the literature [3].



**Fig. 46.3** ROC curve for method of this paper with single detector (*left*). ROC curve for method of this paper with literature [3] (*right*)



**Fig. 46.4** The examples of detections in different environments. **a** Example of detections using local detectors in INRIA. **b** Example of detections using the method in this paper in CAVIAR. **c** Examples of detections in self-timer video

Figure 46.4 shows the test results of this method in different environments. (a), (b) show the test results only using local detectors in test set of INRIA and the video database CAVIAR respectively. And (c) is the test results of the self-timer video image. Where, the upper row of (a) and (b) show the detected pedestrians marked by a rectangle box in the image, and the following row show the parts of a

detected pedestrian marked using different parts detectors in a chosen rectangle box above which had been amplified by a certain percentage. The method of this article can still achieve a desired effect when body parts are occluded. In short, the proposed method based on motion information and part detectors can accurately detect pedestrians in various environments and postures and object occlusion, reflecting the robustness of the proposed method.

## 46.6 Conclusion

The method based on part detectors and motion information using to solve the problem that the target exist occlusion in practical sense is proposed in this paper. The detection speed can be improved by extracting a motion area as candidate region of pedestrians in where pedestrians will be further confirmed. The test results in a different test set show that, the proposed method can be effectively applied to the pedestrian detection task of target occlusion handling. However, the local detectors are trained based on HOG feature pyramid and combined with linear SVM classifier in our method. In the first instance, a large part of time is consumed for HOG calculation. Secondly, the detection rate may be suppressed by the linear SVM classifier which is binary classification. Therefore, the subject of following research is on reducing the dimension of HOG and improving the SVM, to arrive at better detection efficiency.

## References

1. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE computer society conference on computer vision and pattern recognition (CVPR). San Diego, CA, USA, pp 886–893
2. Andriluka M, Roth S, Schiele B (2009) Pictorial structures revisited: people detection and articulated pose estimation. In: IEEE conference on computer vision and pattern recognition (CVPR). Miami, FL, USA, pp 1014–1021
3. Felzenszwalb P, Girshick R, McAllester D et al (2010) Object detection with discriminatively trained part-based models. *IEEE Trans Pattern Anal Mach Intell* 32(9):1627–1645
4. Gao T, Packer B, Koller D (2011) A segmentation-aware object detection model with occlusion handling. In: IEEE conference on computer vision and pattern recognition (CVPR). Colorado Spring, CO, USA, pp 1361–1368
5. Bourdev L, Maji S, Brox T et al (2011) Detection people using mutually consistent poselet activations. *ECCV*, Heraklion
6. Bourdev L, Malik J (2009) Poselets: body part detectors trained using 3d human pose annotations. In: IEEE 12th international conference on computer vision, pp 1365–1372
7. Parikh D, Zitnick CL (2011) Finding the weakest link in person detectors. In: IEEE conference on computer vision and pattern recognition (CVPR). Colorado Spring, CO, USA, pp 1425–1432



8. Anh TN, Auvinet E, Meunier J (2012) Head detection using kinect camera and its application to fall detection. In: 11th international conference on information science, signal processing and their applications (ISSPA). Montreal, QC, Canada, pp. 164–169
9. Bouttefroy P, Bouzerdoum A (2010) On the analysis of background subtraction techniques using gaussian mixture models. In: IEEE international conference on acoustics speech and signal processing (ICASSP), Dallas, TX, Miami, FL, USA, pp 4042–4045
10. Maji S, Malik J (2010) Object detection using a max-margin though transform. In: IEEE computer society conference on computer vision and pattern recognition (CVPR), Miami, FL, USA pp 1038–1045

# Chapter 47

## Tracking Algorithm Based on Joint Features

Xiaofeng Shi and Zhiping Zhou

**Abstract** In the study of target tracking process, when the target has a similar color to the background easily leads to the loss of the target due to illumination and noise. In order to avoid the drawback of Mean shift which only uses color information as the features to track the target, Sobel operator and local binary patterns (LBP) are combined to extract the textures of the moving target as Mean shift characteristics. An advantage of the Mean shift algorithm can compute the histogram easily. However, this process can't change the size of a search window. Therefore, the proposed method extracts the feature points of the object in the region that given by the improved Mean shift and according to the information that the positions of the special feature points, a new search window is generated. The experiments show that the proposed object tracking system performs more accurately than the Mean shift algorithm.

**Keywords** Object tracking · LBP · Sobel operator · Variable search window · Mean shift

### 47.1 Introduction

The target tracking is essentially based on the image frame in the form of the continuous position of the target through the matching features. Currently, moving target tracking is required in many practical applications, such as video surveillance, traffic monitoring to video analysis and understanding [1]. However, the problems of illumination, occlusion and deformation in complex environments remain challenging in the tracking field [2]. A lot of tracking algorithms have been

---

X. Shi (✉) · Z. Zhou  
School of Internet of Things Engineering, Jiangnan University, Wuxi,  
214122 Jiangsu, China  
e-mail: shixiaofengtracy@yeah.net

proposed to overcome the problems. Among them, Mean shift tracking algorithm based on regional representation has recently become popular due to its simplicity and efficiency.

A weighted color histogram is a discrete estimation of the probable density distribution of the target region. It has a good robustness, such as the rotation invariance. However, this method may result in the loss of tracking target when the target has a similar color distribution to the background. As we all know, the color histogram contains no spatial information. In order to solve the above problem, a variety of improved methods have been proposed by scholars in recent years. Shen et al. [3] proposed a kernel-based tracker generated by allowing to build the target model through support vector machine training and to update the target model online. But this algorithm was at the cost of some computational efficiency. In paper [4], texture pattern was applied to the Mean shift algorithm to represent the moving object. Only the main uniform patterns were extracted to reduce the computational complexity. But this process can't change the size of a search window. Yang et al. [5] proposed a tracking approach based on "bag of features". Though this method could track the target accurately, the model updating scheme needed more computational work. SIFT has the advantage scaling and rotation invariant, therefore it can be combined with other tracking algorithms to reduce the complexity. Kuo [6] trained samples online and allowed models to adapt to object instances. In the representation of target's appearance, HOG and RGB features were fused. However, HOG feature is generally for human identification. When it comes to other non-human objects tracking may cause errors.

This article is organized as follows: Sect. 47.2 presents a brief introduction to the Mean shift algorithm and features selection. In Sect. 47.3, our proposed approach is introduced for the target tracking. The experimental results are in Sect. 47.4. Conclusions are given in Sect. 47.5.

## 47.2 Object Tracking

Mean Shift algorithm [7] is popularly used for tasks such as clustering, probability density estimation and tracking because of its simplicity and real-time characteristics. This algorithm is considered as an efficient color space analysis technique and aims at finding local maxima in any probability distribution.

Denote by  $x_i (i = 1, \dots, n)$  the normalized pixel positions in the target region, the target model is computed as

$$q_u = C \sum_{i=1}^n k\left(\|x_i\|^2\right) \delta[b(x_i) - u] \quad u = 1, \dots, m \quad (47.1)$$

where  $q_u$  represent the probabilities of feature  $u$ ,  $m$  is the number of feature spaces,  $\delta$  is the Kronecker delta function,  $b(x_i)$  associate each pixel which is mapped to the corresponding feature spaces.  $k(x)$  is an isotropic kernel profile,  $C$  is a normalization

constant which meets the demand of function  $\sum_{u=1}^m q_u = 1$ . The histogram of the target can be obtained:

$$P_u(y) = C_h \sum_{i=1}^{n_h} k \left( \left\| \frac{y - x_i}{h} \right\|^2 \right) \delta[b(x_i) - u] \quad u = 1, \dots, m \quad (47.2)$$

where  $x_i (i = 1, \dots, n)$  represent the pixel positions in the candidate target region,  $C_h$  is a normalization constant,  $n_h$  is the number of pixels in the candidate target region,  $h$  is the band width. The weight of each pixel is defined as follows:

$$w_i = \sum_{u=1}^m \sqrt{\frac{q_u}{P(y_0)}} \delta[b(x_i) - u]. \quad (47.3)$$

## 47.3 The Proposed Tracking Algorithm

### 47.3.1 Features Selection

#### 47.3.1.1 Local Binary Pattern Combined with Sobel

The traditional tracking algorithm is very effective when the target's color is unique compared with the background's color. However, color information is not enough to describe the object's feature. To solve this problem, we can choose another feature as the object model. In this paper, LBP texture and Sobel edge feature are chosen as the substitute.

LBP is widely used in many fields such as face recognition, texture classification because of its advantage of the gray-scale and rotation invariance [4]. The function of the LBP operator is defined as follows:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (47.4)$$

Where the thresholding function  $s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$ .  $g_c$  represent the gray value

of the center pixel,  $g_p$  corresponds to the gray value of pixels on a circle with radius  $R$ .  $P$  is the number of the local neighbor pixels. Take both speed and discrimination capability consideration, we adopt LBP operator with the parameters  $P = 8$  and  $R = 1$  to describe the texture features in this paper.

$$LBP_{8,1}(x_c, y_c) = \begin{cases} \sum_{p=0}^7 s(g_p - g_c), & U(LBP_{8,1}) \leq 2 \\ 9, & otherwise \end{cases} \quad (47.5)$$

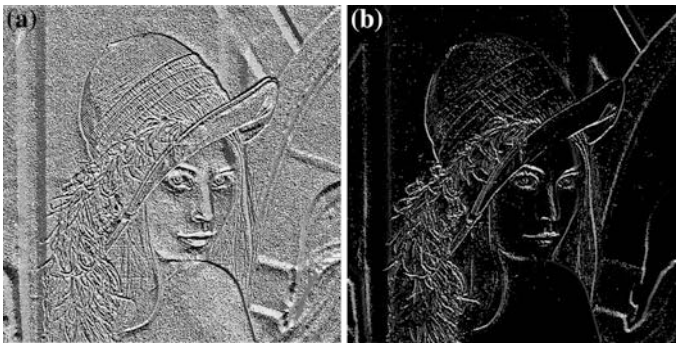
$$U(LBP_{8,1}) = |s(g_7 - g_c) - s(g_0 - g_c)| + \sum_{p=1}^7 |s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (47.6)$$

The  $LBP_{8,1}$  model has nine uniform texture patterns including spots, flat areas, edges, line ends and corners, etc. Three dimensional RGB color channel and one dimensional texture channel can be defined as color-texture histogram. However, the experimental results show that compared with only use of color histogram, this directly define the joint color-texture histogram can't have a strong discrimination ability especially when the background has a similar color with target. In order to improve the tracking effect, interference of the flat area should be reduced while the features of uneven region should be enhanced. Pattern 2, 3, 4, 5, 6 represent the main features of the target, and the others are minor textures. Thus, the main uniform patterns of the target can be defined by the following equation:

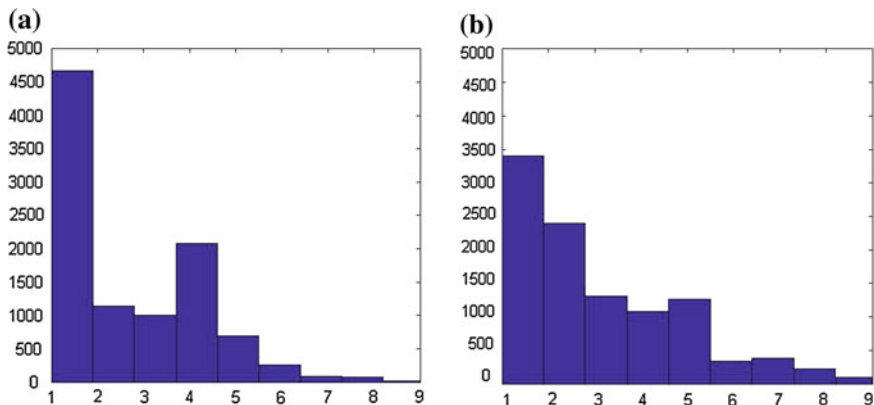
$$FLBP_{8,1}(x_c, y_c) = \begin{cases} \sum_{p=0}^7 s(g_p - g_c), & U(LBP_{8,1}) \leq 2 \text{ and } \sum_{p=0}^7 s(g_p - g_c) \in \{2, 3, 4, 5, 6\} \\ 0, & otherwise \end{cases} \quad (47.7)$$

As shown in Fig. 47.1, such texture feature only considering the vigorous pixels of the target area can ignore the flat elements of the background and the target area, which can make up the deficiency of color histogram.

Actually, LBP operator encodes the information of edge, corner and other local information in an image. In order to extract the more detailed information, Sobel operator is used in this paper due to its low computational complexity. Figure 47.2



**Fig. 47.1** Texture on Lena: **a** Texture extracted by LBP **b** Texture extracted by the five LBP major uniform patterns



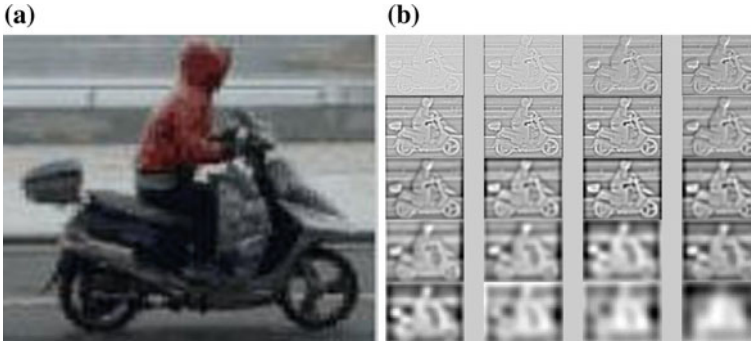
**Fig. 47.2** Nine patterns distribution: **a** LBP's nine patterns distribution, **b** LBP's nine patterns distribution combined with Sobel

shows the nine patterns distribution in the ROI of Fig. 47.3a. Compared to Fig. 47.2a, b has more detailed information, such as Pattern 2, 3, 4, 5, 6. Sobel operator [8] is so efficient to implement that slightly increases the computational work in process of feature extraction by combining Sobel operator with LBP. Sobel operator has two convolution kernels. The one kernel is usually corresponding to the largest vertical edge, and the other corresponding to the largest horizontal edge. The LBP-Sobel operator can be defined as follows:

$$\begin{cases} I'_x = \sum_{p=0}^{p=7} s(I_x - I'_c)U(LBP_{8,1}) \leq 2 \text{ and } \sum_{p=0}^7 s(g_p - g_c) \in \{2, 3, 4, 5, 6\} \\ I'_y = \sum_{p=0}^{p=7} s(I_y - I'_c)U(LBP_{8,1}) \leq 2 \text{ and } \sum_{p=0}^7 s(g_p - g_c) \in \{2, 3, 4, 5, 6\} \end{cases} \quad (47.8)$$

Where  $I_x$  and  $I_y$  represent the horizontal and vertical edge detection image respectively.

The experiments show that representing the target by using color and the main uniform patterns of LBP-Sobel texture on the whole tracking region is hard to track the object because of the strong interference from the background. Therefore, a new way is to combine the color and LBP-Sobel texture features. For example, color space is quantified for  $16 \times 16 \times 16$  to describe the target in Mean shift algorithm. When the texture edge features are added, color features can be rough for the quantification for  $8 \times 8$  that is R and G two color channels, In addition to five LBP-Sobel models, so only  $8 \times 8 \times 5$  can meet the demand. Obviously, the vector number of space quantization is less than color histogram.



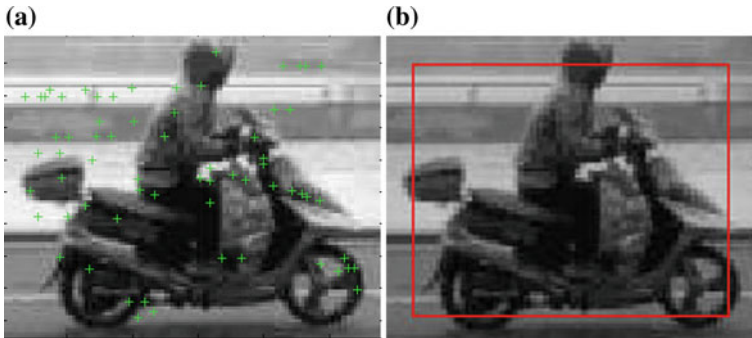
**Fig. 47.3** Gaussian pyramids of ROI: **a** ROI, **b** Gaussian pyramid

### 47.3.1.2 SIFT Algorithm

The SIFT algorithm was introduced by David G. Lowe, which is used in feature extraction and matching. The process of this algorithm can be divided into two major aspects: the detector and descriptor categories. It generally has four steps [9, 10]. Firstly, construct a Gaussian pyramid and search for local peaks in a series of difference of Gaussian (DoG) images. For each octave of scale space, the initial image is repeatedly convolved with Gaussian to produce the set of scale space images shown in Fig. 47.3b. Secondly, key points are localized and removed if found to be unstable. Thirdly, identifies the dominant orientation for each feature point based on its local image patch. The assigned orientations scale and location for each feature point enable SIFT to construct a canonical view for the feature points those are invariant to similarity transforms. The final stage builds a local image descriptor for each feature point, based upon the image gradients in its local neighborhood.

### 47.3.2 The Process of Tracking Algorithm

Mean shift algorithm can compute the histogram easily, but this process can't change the size of a search window. Camshift algorithm can change the search window. However, it is sensitive when the moving object's color is similar to the background's color due to illumination and noise. The weakness may restrict the tracking result. Therefore, to overcome this weakness, in this article, we present the improved tracking algorithm. Firstly, combine the Mean shift with the five LBP major uniform patterns to get the rough location of the targets. Secondly, detect the feature points by using SIFT in the fixed search window. Thirdly, select the special feature points. These feature points should meet the requirements: (1) The maximum value of the abscissa and ordinate. (2) The least value of the abscissa and ordinate. (3) The maximum value of the abscissa and the least value



**Fig. 47.4** An example of the generation of the new search window: **a** Feature points using the SIFT algorithm, **b** The new search window

of the ordinate. (4) The least value of the abscissa and the maximum value of the ordinate. Finally, according to the special feature points, we can get a new rectangle. Compared with the rough location, this algorithm can locate the target more accurately. Figure 47.4a shows the feature points by using the SIFT algorithm in ROI. Figure 47.4b shows the new search window.

## 47.4 Experimental Results

This experiment is implemented using Matlab R2009b programs and carried out on a PC with a T5250 1.50 GHz processor with 1.00 G of memory. In the experiment, the video sequence that has a total of 251 frames, a size of  $640 \times 480$  is captured for a casual shot.

From Table 47.1, we can see that Mean shift simply needs to quantify the color. When combined the LBP-Sobel operator, its computational work is more complex because it involves in the points of adjacent pixels. However, in the tracking process, time is mainly spent on iterative calculation. Compared with the Mean shift based on all points, the improved method needs 41.8 % of points involved in the calculation. The localization errors of the two methods are calculated (refer to Table 47.2). The Table shows that the proposed method can track the object more accurately and it achieves the smaller mean and standard deviation values between the two tracking algorithm.

**Table 47.1** Comparison of numbers of mean shift iterations of the two algorithms

	Color	Fusion
Total number of iteration	816	557
Average number of iteration/(frame)	3.252	2.219



**Table 47.2** Localization error

	Mean shift based on LBP-Sobel	Proposed method
Mean error/pixel	7.12	2.89
Standard deviation/pixel	8.47	3.26

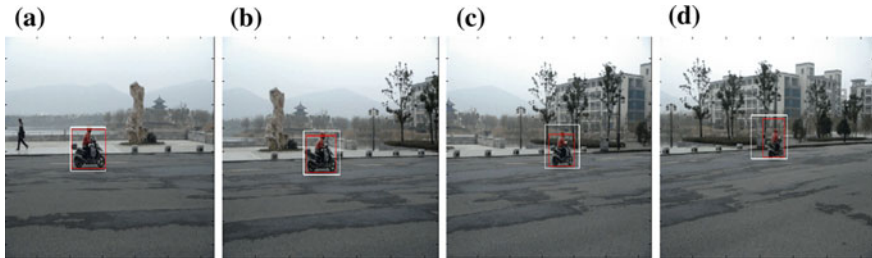
**Fig. 47.5** Comparison of the resulting images of object tracking using different algorithms: **a** frame 38, **b** frame 64, **c** frame 96, **d** frame 124

Figure 47.5 shows the comparison with resulting images of object tracking using different algorithms. A red rectangle denotes the proposed method, a white rectangle indicates the Mean shift algorithm based on LBP and Sobel operator. The proposed method can track the object more accurately than another algorithm. Moreover, it is clear that the search window is adapted to the size of the detected object for the tracking with the proposed method. Compared with two results, we can see that when the target is far from the camera, size of the target is getting smaller and smaller. The red rectangle changes from small to big itself according to the variable size. Difference of the two experimental results is more and more obvious with changes of target's size.

## 47.5 Conclusions

A joint color and the main uniform patterns of LBP texture to present target model are useful in suppressing the smooth background interference. These features based Mean shift tracking algorithm is proposed to get the rough location of the targets. SIFT feature detection is used to reducing the size of the search window aimed at tracking the target more precisely. This method has demonstrated that with the change of the object size, the search window size will change, too. This paper makes up for the deficiencies of the traditional tracking algorithm. But how to track the multiple targets when they are completely occluded in complex background needs a further study.

## References

1. Yin HP, Chai Y, Yang SX et al (2011) An improved mean-shift tracking algorithm based on adaptive multiple feature fusion. *Informatics in control automation and robotics*, Milan pp 49–62
2. Li GB, Wu HF (2011) Scale adaptive ensemble tracking. 4th international congress on image and signal processing, Shanghai, pp 431–435
3. Shen CH, Kim J, Wang HZ (2010) Generalized kernel-based visual tracking. *IEEE Trans Circuits Syst Video Technol* 20(1):119–130
4. Ning JF, Zhang L, Zhang D et al (2009) Robust object tracking using joint color-texture histogram. *Int J Pattern Recognit Artif Intell* 23(7):1245–1263
5. Yang F, Lu H, Chen YW (2010) Bag of features tracking. *International conference on pattern recognition 2010*, Istanbul, pp 153–156
6. Kuo CH, Huang C, Nevatia R (2010) Multi-target tracking by on-line learned discriminative appearance models. *IEEE conference on computer vision and pattern recognition 2010*, San Francisco, pp 685–692
7. Comaniciu D, Meer P (2002) Meanshift: a robust approach toward feature space analysis. *IEEE Trans Pattern Analysis Mach Intell* 24(5):603–619
8. Zhao S, Gao Y, Zhang B (2008) Sobel-LBP. *IEEE international conference on image processing*, San Diego, pp 2144–2147
9. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vision* 60(2):91–110
10. Ha S-W, Moon Y-H (2011) Multiple object tracking using SIFT features and location matching. *Int J Smart Home* 5(4):17–26

# Chapter 48

## Distributed Audit Secure Data Aggregation for Wireless Sensor Networks

Zhengdao Zhang and Zhiping Zhou

**Abstract** Data aggregation can reduce the communication overhead and energy expenditure of sensor nodes, as well as extend the life-cycle of the wireless sensor network. However, because individual sensors may be compromised, the data aggregation also introduces some risks including the false data injection attacks. This paper proposes a distributed audit secure data aggregation protocol. The aggregates are audited at the next level nodes of the aggregators. The communication overload, which Base Station (BS) originates in the attest process, can be avoided. Furthermore, because we can find the false data in the lower level, it is easier to strike out the false data, and only a little fraction of readings are dropped off. To do these, the aggregators attach multi-certificates to the aggregates. Those certificates may include the maximum, minimum, mean reads and those nodes' identifiers. To further reduce the communication overload, we use the watermark method to embed the multi-certificates in authentication part of aggregates. The length of message is kept as same as that under the normal hop-by-hop aggregation protocol with MACs. The analysis shows that our protocol is efficient and provides certain assurance on the trustworthiness of the aggregation result.

**Keywords** Data aggregation · Distribution · Probabilistic grouping · Watermark · Sensor network security

### 48.1 Introduction

Wireless sensor networks are usually consist of hundreds or thousands of inexpensive, low-powered sensing devices with limited memory, computational, and communication resources [1]. Due to the low deployment cost requirement of

---

Z. Zhang (✉) · Z. Zhou  
Engineering Research Centre of IOTs Technology and Application, Ministry of Education,  
Jiangnan University, Wuxi, Jiangsu, China  
e-mail: wxzdz.dr@gmail.com

wireless sensor networks, sensors only have limited computation and communication resources. To reduce the communication cost and energy expenditure in data collection, many data aggregation protocols have been proposed. The data aggregation reduces the number of data transmissions, thereby improving the bandwidth and energy utilization in the network. On the other hand, because the raw data will be aggregated at aggregators, data aggregation introduces a lot of security vulnerabilities. For example, the BS cannot authenticate the raw data and find the compromised nodes in network. More seriously, the compromised aggregators can inject an arbitrary false fusion result to make the final aggregation result to far deviate from the true measurement. A lot of researchers focus on the secure data aggregation protocol [2, 3], such as SDA [4], SDAP [5], RSDA [6] and so on. But, the existing methods have some common drawbacks. First, all of them need the BS to attest all suspicious, that introduces a heavy communication overload. Second, if an aggregation result is verified as an attack, the BS drops off this value as well as all reads between attack injection point and BS. This introduces an extra waste of sensor ability.

To solve these problems, a distributed audit secure data aggregation protocol based on the principles of a distributed and cooperative attestation is proposed. A random grouping technique is used to construct a tree topology and logical groups of nodes. The leader node generates an aggregation result. The sibling nodes and the parent node of that leader identify the suspicious reads based on the set of results and a summary of historic reads. After that, the suspicious node in that attestation process proves the correctness of its read. Finally, each group aggregate the result which passes attestation reply to the upper leader and participates in the next aggregation until the result reaches the BS. The remainder of this paper is organized as follows. Section 48.2 describes our data aggregation protocol composed of network model, grouping, aggregation and audit. Security analysis and performance evaluation of our scheme are presented in Sect. 48.3. After that, we summarize our work in Sect. 48.4.

## 48.2 The Security Data Aggregation Protocol

We assume the BS has unlimited energy, and cannot be compromised. Also, there is a topological tree rooted at the BS which the shape and the distance (in number of hops) between every node are unknown. The transmission mechanism in the protocol is reliable. We also assume there is a broadcast protocol which can provide global broadcast authentication. Every sensor node shares their individual public key with each other. The aggregation tree is constructed as [5]. The count number of  $y$  is named  $C_y$ .

In our method, the credentials including IDs and Readings will be embedded into a cover message which is calculated based on the message authentication code (MAC) of sensitive data. These credentials will be extracted in the audit procedure by other nodes. The credentials have the format like  $(ID_i|R_i)$  The cover messages

always like  $MAC(K, ID_i|flag|R_i)$  which is a MAC calculated with the key  $K$ . Let  $len(\cdot)$  denote the length of data. The total bits of sensitive elements should be less than that of cover data,  $\sum len(ID_i|R_i) < MAC(K, ID_i|flag|R_i)$ . The details of algorithms can be found in [7].

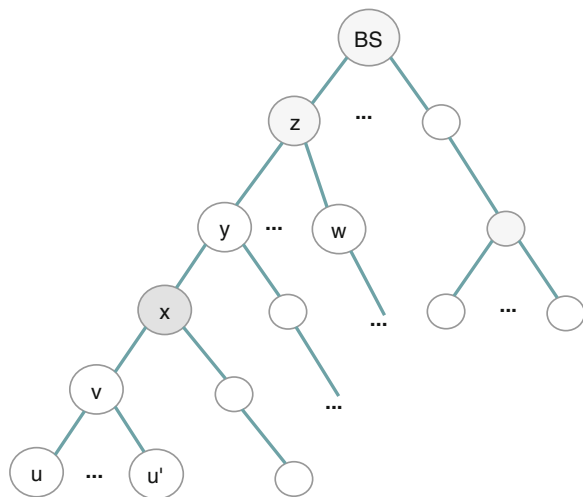
### 48.2.1 Subsection Query Dissemination and Leader Selection

We choose the leader nodes as the SDAP [5] does. An example of an aggregation tree and the result of probabilistic grouping are showed in Fig. 48.1. BS first decides the aggregation function  $F_{agg}$  and randomly chooses a number  $K_{S,BS}$ . The BS broadcasts a packet as  $BS \rightarrow \rightarrow * : ID_{BS}, K_{S,BS}, F_{agg}$ . When a leader node  $x$  receives query packet from its parent node  $y$ ,  $x$  stores the  $F_{agg}$  and  $K_{S,Z}$ . After that,  $x$  uses a one-way function  $H$  to calculate the key for its group as  $K_{S,X} = H(K_{S,Z}, ID_X \bmod C_X)$ . So,  $x$  broadcasts query packet to its children nodes along the tree as  $X \rightarrow \rightarrow * : ID_X, K_{S,X}, F_{agg}$ . When an intermediate node receives a query from its parent node, it stores  $F_{agg}$  and  $K_{S,X}$  of the group and broadcasts the query in same way [8].

### 48.2.2 Aggregation Commitment

Each aggregation packet contains two parts. One is the packet header contained a sender's ID and a flag which implicates different duty of the packet. Another is the data part which contains node's IDs and node's sensed data. We use MAC as a

Fig. 48.1 An example of the aggregation tree



cover message, and we embed the node's IDs and the node's sensed data in it [9]. In the left part, we use  $WM(key, cover\ message, hiding\ data)$  to denote the data part.

The aggregation process starts from the leaf nodes towards the BS. The leaf node sends its ID, data to its parent. The packet sent to  $v$  from  $u$  is:

$$u \rightarrow v : ID_u, flag = 0, C_u = 1, WM(K_{S,X}, MAC(K_{S,X}, ID_u | flag | R_u), (ID_u | R_u))$$

where the  $R_u$  is the sensed data of node  $u$ . The flag value '0' means this packet doesn't need to be audited,  $C_u = 1$  means there is only one data part, and  $MAC(K_{S,X}, ID_u | flag | R_u)$  means the MAC value which is computed with the key  $K_{S,X}$ .

When a leader node, named  $x$ , receives a packet, it first checks whether this packet comes from its child or not. If not, it just drops it. Otherwise,  $x$  uses the group key  $K_{S,X}$  to extract the hidden data from the data part of the packet. Then,  $x$  calculates the MAC with  $K_{S,X}$  again, and compares it with the cover message. The packet is accepted only if the authenticity of the MAC is successful.

After obtaining all children's readings  $R_1, R_2, \dots, R_{C_x-1}$ ,  $x$  first rearranges all of them with the reading  $R_x$  from  $x$  in the ascending order of value. Then,  $x$  aggregates them as  $R_{aggx} = F_{agg}(R_1, R_2, \dots, R_{C_x})$ , and uses the key of the upper group to send upward to its parent, named  $y$  in Fig. 48.1.

$$x \rightarrow y : ID_x, flag = 1, C_x, WM_1, WM_2$$

$$WM_1 = WM(K_{S,Z}, MAC(K_{S,Z}, ID_1 | flag | R_1), (D_1, R_1)), \dots, \\ WM(K_{S,Z}, ID_{C_x} | flag | R_{C_x}), (ID_{C_x} | R_{C_x})$$

$$WM_2 = WM(K_{S,Z}, MAC(K_{S,Z}, ID_x | flag | R_{aggx}), (D_x, R_{aggx})).$$

here, flag value '1' means that the values in this packet need to be audited.

When an intermediate node receives a packet from its child node, it then checks the flag. If the flag is '0', it directly forwards the packet to its parent node along the logic tree. Otherwise, the intermediate node will perform an audit process to verify the correctness of the aggregation value, find the suspicious data, and attest the suspicious data. If all individual sensing data pass the verification, the intermediate node sets the flag value '0', sets the count number  $C_x = 1$  again, deletes all individual cover messages ( $WM_2$  in data part) and only keeps the last cover message embedded with the aggregation data. Then the intermediate node sends this new packet upward to its parent. For example, in Fig. 48.1,  $y$  sends the packet to its parent as follow, if the verification process is successful.

$$y \rightarrow * : ID_x, flag = 0, C_x = 1, WM(K_{S,Z}, MAC(K_{S,Z}, ID_x | flag | R_{aggx}), (D_x, R_{aggx})).$$

### 48.2.3 Data Verification and Audit

After a parent node of the group leader, say  $y$ , has received an aggregation messages, it needs to verify the authenticity of the aggregated value. This includes verifying the content of the packet and the authenticity of the leader. First, based on the  $K_{S,Z}$ ,  $y$  can extract hiding data from each cover message and get the flag and the count value  $C_x$ . Secondly, the BS verifies the count number of leader falls in certain range or not. Based on our group partition way, the count number is no more than  $k * k$ , where  $k$  is the maximum number of a logic group. If the count number is large, we have more reasons to suspect that the leader is performing count changing attack. So, the probability for individual audit should be increased.

Choose two cluster centers  $B_1, B_2$  and  $B_0 = (B_1 + B_2)/2$  as three references. Let  $r = (B_2 - B_1)/4$  as a radius and count the reading which falls in the different regions around them. If no attack is lunched, these references are similar. Otherwise, if the count number of the region around  $B_0$  is much smaller than others, we can suspect one of clusters is forged. If the adverse injects the false readings that gradually deviate from the normal value, the distance between two centers  $d = (B_2 - B_1)$  is large. We also can find the suspects easily.

In two situations, the node will ask BS to help perform an audit. One is the parent of a leader suspects that there are some nodes were compromised; another is sibling nodes of leader decide to audit some individual nodes with a per-set probability  $P_C$ . To the first situation, the node that has the minimum reading or maximum reading will be audited at first. Then, if anyone fails the audit, all nodes whose reading falls into the same cluster with failed node will be audited. The parent node  $y$  sends an audit request to  $ID_1$  with the public key  $K'_{ID_1}$  of  $ID_i$  as

$$y \rightarrow ID_i : ID_y, flag = 3, WM\left(K'_{ID_1}, MAC\left(K'_{ID_1}, flag|ID_y|K'_y\right), \left(ID_y|K'_y\right)\right)$$

$ID_i$  then uses  $y$ 's public key  $K'_y$  to response this request as

$$ID_i \rightarrow y : ID_i, flag = 3, WM\left(K'_y, MAC\left(K'_y, flag|ID_i|R_i\right), (ID_i|R_i)\right)$$

After receiving  $R_i$ ,  $y$  compares it with the reading extracted from the leader. If these two readings are same, the node  $ID_1$  passes the audit. These readings will be transmitted to the group leader as an individual aggregation result. They will be accepted only other readings or aggregation results in this group can support it. If two readings are different, we know the leader or the  $ID_i$  was compromised. So,  $y$  will report the compromise to its group leader  $z$ . Supposed that  $ID_1$  was failed in audit.  $Y$  recalculates the aggregation result as  $R'_{aggx} = F_{agg}(R_1, \dots, R_{i-1}, R_{i+1}, \dots, R_{Cx})$ . Then,  $Y$  reports the compromise with the new result.

$$y \rightarrow z : ID_y, flag = 0, C_x = 1, WM\left(K_{sg,z}, MAC\left(K_{sg,z}, ID_x|flag|C_x|R'_{aggx}\right), \left(ID_x|R'_{aggx}\right)\right), \\ WM\left(K_{sg,z}, MAC\left(K_{sg,z}, ID_i|R_i\right), (ID_i|R)\right)$$

The sibling node first extracts all readings from the cover messages with the group key. Then, for each reading, sibling node performs an audition procedure based on a per-set probability named  $P_C$ . So, the sibling node uses a random number generator to get a random number. If the number is less than  $P_C$ , the current reading will be audited as above. This process will repeat until all readings take part in the audit process. After that, if any conflicts are found, the sibling node also reports them to the group leader. In Fig. 48.1, node  $w$  acts as a sibling node and it will report the compromise as:

$$w \rightarrow z : ID_w, flag = 4, C_p WM(K_{sgg,z}, MAC(K_{sg,z}, flag | ID_w | R_w), (ID_w | R_w)).$$

### 48.3 Security Analyses and Performance

In our method, BS always disseminates aggregation query with a different random number  $K_{s,SB}$  and the number acts as a key as well. Based on this original key, every group produces a different key for every round too. If the malicious node replies past reading again, the leader node and the other sibling nodes can't extract the correct hiding data with a new group key and will get a different MAC. Then, this packet will fail in the authentication and be discarded.

An attacker cannot selectively compromise some nodes and makes some of them to appear in same group, because the groups are decided by a random process. A compromised node may change a real value corresponding to an abnormal event to a normal value. This attack, named potential event suppression attack, is a big vulnerability to the data aggregation. Our method can deal with this attack because the sibling nodes of leader will audit every reading with a probability. If the leader gets rid of the real abnormal node from its children list, the parent node can easily find that by checking the leader selection condition. Our method mitigates the count changing attack because (1) every reading taking part in aggregation will be audited with same probability, and (2) the  $ID$  of every node is unique and connected with a pair of keys which is stored at BS.

In the group, every leaf node only can report a reading with count number '1'. So, if the attacker launches a count changing attack, the compromised node must be a group leader node. Then, to the each sibling node of that leader node, it will audit every reading with a per-set probability  $P_C$ . If the compromised node forges a real  $ID$  and the compromised node can't get the key pair of that  $ID$ , the sibling node asks for audit information, the real node responses that query. Then the sibling node will easily find the attack because the two readings are different.

A compromised node also may forge small count but extreme data to modify the final aggregation result. In this situation, the attacker also needs to compromise a group leader node, and then forge the readings of its children nodes. This is as same as the second case of count changing attack, and can be detected as the same way.



For simplicity, we only consider the case where the leaf nodes transmit their readings and no readings are expected from aggregator nodes. We assume a general tree hierarchy in which every node has  $b$  children and the depth of the tree is  $d$  as same as in [2]. Assume the length in bits of reading from the leaf nodes is  $x$ . The sensor node ID in bits will be denoted as  $y$ . Let the bits of count number equals 8. Also, we denote the MAC's length in bits as  $z$ . Since TinyOS packet has the maximum size of 36 byte, including 29 byte payload and 6 byte header, we denote header as  $oh$  to compute the overhead bits transmitted within the network.

We assume each group has an average size of  $s$ , so the number of the groups is  $(N/s) + 1$ . Also, the distance from each leader to the BS can be considered as  $d/2$ . If a node sends its  $ID$ , three bits of flag, a count number and a MAC, the length is  $y + 3+8 + z+oh$ . The total number of bits in aggregation is approximated by

$$\left(\frac{b^d - b}{b - 1}\right)(s - 1)(y + z + 11 + oh) + \left(\frac{N}{s}\right)(s - 1)z$$

And, the overhead for attestation is around

$$\left(\frac{b^{d-1} - b}{b - 1}\right)\left(2(s - 2)\left(\frac{N}{s}\right)\left(\frac{d}{2}\right) + 1\right)(y + z + 3 + oh)$$

Our protocol has the same communication overload in aggregation procedure as SDAP does. In the attestation procedure, our protocol has fewer communication overloads if there are several attested groups. Based on the same simulation assumption with [5], every attestation procedure of SDAP need relay no less than  $88 \text{ packet} \cdot \text{hop}$ . If there are  $n$  attested groups, the relay is no less than  $88 \cdot n \text{ packets} \cdot \text{hop}$ . Our protocol totally need relay  $95 \cdot 30 \cdot P_c \text{ packets} \cdot \text{hop}$  for any number of suspect groups. Notice that the  $P_c$  is a small number in our protocol, E.g. 0.1. Then, our protocol relay fewer packet $\cdot$ hop if the number of attested group is more than 3. If the number of attested group is 10, our protocol only need relay  $32.39 \% \text{ packet} \cdot \text{hop}$  of SDAP.

## 48.4 Conclusion

In this paper, we propose a Distributed Audit Secure Data Aggregation Protocol for large-scale sensor networks. By using the probabilistic grouping method, we partition the aggregation tree into the different logic groups. The upper level nodes can also check the grouping results to prevent some malicious nodes to compromise the special group leaders. A cluster procedure is performed at aggregator to find any suspect readings, and an audit procedure is performed in the next upper group of aggregator to audit the aggregation result. To simplify the audit process, we attach multi-certificate in the aggregates, and use watermark algorithm to keep the message length no longer than the message length in other protocol. In the future, the potential of integrating with CDMA watermark algorithm will be investigated too.

**Acknowledgments** This work is supported by Fundamental Research Funds for Central Universities, Chinese Department of Education, under Grant JUSRP111A49.

## References

1. Yick J, Mukherjee B, Ghosal D (2008) Wireless sensor network survey. *Comput Netw* 52(12):2292–2330
2. Hani A, Ernest F, Juan M, Gonzalez N (2008). Secure data aggregation in wireless sensor network: a survey. In: 6th Australasian information security conference, vol 81 of CRPIT, Wollongong, NSW, Australia, pp 93–105
3. Ozdemir S, Xiao Y (2009) Secure data aggregation in wireless sensor networks: a comprehensive overview. *Comput Netw* 53(12):2022–2037
4. Hu L, Evans D (2003) Secure aggregation for wireless network. In: SAINT workshops, IEEE Computer Society, pp 384–394
5. Yang Y, Wang X, Zhu S et al (2008) SDAP: a secure hop-by-hop data aggregation protocol for sensor networks. *ACM Trans Inf Syst Secur* 11(4):1–43
6. Alzaid H, Foo E, Nieto G (2008) RSDA: Reputation-based secure data aggregation in wireless sensor networks. In: Proceedings of the 1st international workshop on sensor networks and ambient intelligence, Dunedin, New Zealand
7. Yang J, Sun X, Wang B, Xiao X (2010) Bloom filter-based data hiding algorithm in wireless sensor networks. In: The 5th international conference on future information technology, Busan, Korea, pp 1–6
8. Perrig A, Szewczyk R, Wen V, Culler D, Tygar J (2001) SPINS: security protocols for sensor networks. In: *Mobile computing and networking*, pp 189–199
9. Zhang W, Liu Y, Das SK, De P (2008) Secure data aggregation in wireless sensor networks: a watermark based authentication supportive approach. *Pervasive Mob Comput* 4(5):658–680

# Chapter 49

## Multi-Expression Based Gene Expression Programming

Wei Deng, Pei He and Zhi Huang

**Abstract** Among the variants of GP, GEP stands out for its simplicity of encoding method and MEP catches our attention for its multi-expression capability. In this paper, a novel GP variant-MGEP (Multi-expression based Gene Expression Programming) is proposed to combine these two approaches. The new method preserves the GEP structure, however unlike the traditional GEP, its genes, like those of MEP, can be disassembled into many expressions. Therefore in MGEP, the traditional GEP gene can contain multiple solutions for a problem. The experimental result shows the MGEP is more effective than the traditional GEP and MEP in solving problems.

**Keywords** Genetic programming · Gene expression programming · Multi expression programming

### 49.1 Introduction

Genetic programming is a new branch of the evolutionary algorithm, early GP is organized with structures of tree-based, and operated directly on the tree, which leads to the low efficiency of function mining. To solve this problem, linear genetic encoding and graphic genetic encoding methods have been proposed, such as linear genetic programming (LGP), multi expression programming (MEP) and

---

W. Deng (✉) · P. He · Z. Huang  
School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410114, China  
e-mail: 76995958@qq.com

P. He  
Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China

gene expression programming (GEP). These algorithms nowadays have been applied in a wide range of real-world problems.

Genetic Expression Programming (GEP) algorithm is a new member of the genetic family, which is widely used in knowledge discovery. GEP integrating the GA and GP advantages uses simple encoding method to solve complex problems. GEP still exist some problems that evolutionary generation is too large or unable to obtain optimal results. MEP (Multi Expression Programming) is a GP variant. A unique MEP feature is its ability of encoding multiple solutions for a problem [1]. But the complexity of the MEP decoding process is higher than the other linear genetic programming. In this paper, we propose a novel algorithm MGEP which combines GEP algorithm with MEP (Multi Expression Programming). MGEP employs multi expression features used in GEP. The improved algorithm can achieve higher efficiency of function mining.

This paper realized a new algorithm and related genetic operator, and analyzed the advantages of MGEP algorithm in the usage of expression space compared with the traditional GEP. The algorithm achieved good results for solving specific problems. The Experimental results show that when compared with the standard GEP, the average evolution generation of MGEP is 1.1–5.3 % of the traditional GEP and 7–18 % of MEP.

The paper is organized as follows. The backgrounds of the GEP and the MEP are presented in Sect. 49.2. The MGEP algorithm is described in Sect. 49.3. In Sect. 49.4, we will perform several numerical experiments to compare the MGEP, the GEP and the MEP algorithm. Section 49.5 concludes the paper.

## 49.2 Background

### 49.2.1 MEP and GEP

GEP and MEP are automated method for creating computer programs. GEP as a linear chromosome is composed of genes containing terminal and function symbols. Search operators are operated on a GEP gene to produce new individuals, such as crossover, mutation and transposition. MEP chromosomes are composed of the number of the genes which are substrings of variable length. A unique MEP feature is the ability of choosing the best gene provide the output for the chromosome. This is different from other GP techniques which employed a fixed gene for output. In MEP the offspring obtained by crossover and mutation. These two techniques of the GP variant have been applied in various fields.

**GEP Representation:** A GEP gene is composed of head and tail. The head contains function or terminal symbols, but the tail can only contain terminal symbols. For each problem, the length of the head (denoted  $h$ ) is chosen by the user, and the length of the tail (denoted  $t$ ) is evaluated by formula 49.1, where  $n$  is the maximum number of the function arguments in the function set.

$$t = h * (n - 1) + 1 \tag{49.1}$$

Let us consider a set of function symbols  $F = \{+, *, /, S\}$  and a set of terminal symbols  $T = \{a, b\}$ , where the symbol  $S$  stands for the  $\sin$  operator. In this case  $n = 2$ . If we choose  $h = 7$ , then we get  $t = 8$ , so the length of the gene is  $7 + 8 = 15$ . Such a gene is given below:  $C = +/ * aSb + bbababab$ . GEP is encoded by a fixed length string called K-expression, and the individual phenotype is an expression tree (ET). A K-expression can be translated into an expression tree (ET) by breadth-first parsing. The ET based on the string  $C$  is created in Fig. 49.1. The expression encoded by the gene  $C$  is:  $E_1 = a/\sin(b) + b * (b + a)$ . The last five elements of this gene are not used. Usually, a GEP gene is not entirely used for phenotypic transcription.

**MEP Representation:** A MEP gene encodes a terminal or a function symbol. If a gene contains a function symbol, its arguments must point to gene indices of lower values than the position of the function itself. And the first symbol of the chromosome must be a terminal symbol.

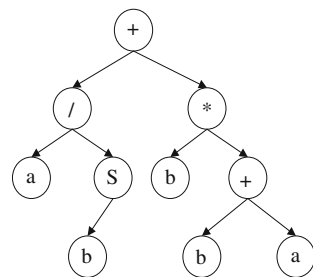
Consider the set of functions:  $F = \{*, +, S\}$ , and the set of terminals  $T = \{a, b\}$ . The numbers on the left stand for gene index and the right is the encoded gene. An example of a chromosome  $C$  using the sets  $F$  and  $T$  is given in Fig. 49.2a. Translation of the MEP chromosome into expressions is done top-down. A terminal symbol decodes a simple expression. A function symbol translates into a complex expression by connecting the operands with the gene indices which would be substituted for the decoded expression. The chromosome in the previous example would be decoded into expressions in Fig. 49.2b.

### 49.2.2 MEP and GEP Related Literature

GEP as a new genetic algorithm shows its superior performance in many fields, especially the outstanding performance in the area of data mining. Nowadays, more and more researches are focusing on GEP all over the world, mainly including the basic principle of GEP and its application.

Some foreign literature reviewed genetic programming [2–5]. We are reviewing the main GP variant with linear representation in paper [2], which provides a

**Fig. 49.1** An expression tree encoding the function  $a/\sin(b) + b * (b + a)$



**Fig. 49.2** MEP encoding and decoding way of the gene  
C a. b

(a)	(b)
1: b	$E_1: b$
2: $\sin(1)$	$E_2: \sin(b)$
3: a	$E_3: a$
4: /3,2	$E_4: a/\sin(b)$
5: +1,3	$E_5: b+a$
6: *1,5	$E_6: b*(b+a)$
7: +4,6	$E_7: a/\sin(b)+b*(b+a)$

complete description for each method. In paper [3] the author outlines some of the challenges and open issues that face researchers and practitioners of GP. The first article of GEP was published by Portugal Candida Ferreira in 2001. Then she wrote many papers and published GEP monographs [6–9]. In paper [9] a hybrid evolutionary technique is proposed for data mining tasks, which combines the Clone Selection Principle with GEP.

MEP has a flexible encoding method and the ability to provide abundant representations. MEP was reviewed systematically in literature [9–11]. All aspects of the MEP are summarized in paper [9]. MEP is widely used in many fields [12, 13]. In paper [12] authors proposed a way of structuring a CGP algorithm to make use of the multiple phenotypes which are implicitly encoded in a genome string.

### 49.2.3 MEP and GEP Weakness

Generally, the more genes in a GEP chromosome, the higher success rate of GEP. However, if the genes are over a certain value, the success rate will decrease. This situation occurs in that GEP uses a complex chromosome to encode a less complex expression. If the target expression is short and the head length is large, the majority of individual is unused. This problem usually arises in evolutionary process when GEP employs chromosome with a fixed length.

There are problems where the complexity of the MEP decoding process is higher than the complexity of the GE, GEP, and LGP decoding processes. This situation usually arises when the set of training data is not a priori known [2].

Consider a GEP gene, as long as we read a K expression from the different head symbols, we can build more expression trees. Owing to the breadth and depth of the same symbols in other expression tree are different, thus the corresponding expressions are not only decomposed by standard GEP expression. Based on this idea, we proposed a novel technique MGEP which combines the advantages of the GEP with the MEP features. The new algorithm would build up more expression trees for a GEP gene. And compared with traditional MEP, MGEP has a diversity of the gene. The MGEP with these two characteristics would overcome the weakness of MEP and GEP.

## 49.3 MGEP Model

### 49.3.1 MGEP Initialization

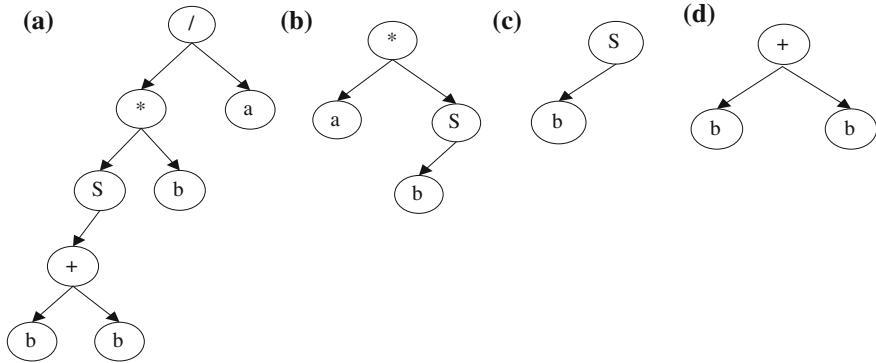
MGEP initialization is the same as the GEP. MGEP chromosomes are initialized randomly, but they must obey two rules that a gene contains two parts of the symbols and the tail of the gene must be composed of terminal symbols.

### 49.3.2 MGEP Representation

The evolutionary process of the MGEP algorithm is similar to the traditional genetic algorithm, namely: initial population → selection → crossover → variation → new population → selection. MGEP encoding way is identical with the standard GEP. But the novel algorithm would decode into multiple sub-expressions by a K-expression. MGEP algorithm as the standard GEP algorithm starts by creating a random population of individuals. To realize our aim, we start by reading the K-expression from the first head character to build the first expression tree, which is totally the same with the GEP. Then, we continue reading the K-expression from the second head character to establish another expression tree. Repeat reading the head characters until the end of the head. Don't consider the single node of the sub-expression tree which called the dead gene.

A MGEP gene contains multiple solutions for a problem. Compared with the MEP algorithm, MGEP have more diverse individuals in the population. Because several MEP genes are disjoined or combined by the sub-expressions, but MGEP gene can decode into diversity of genes. Consider the referred gene C in the second section. We utilize the MGEP algorithm to build multiple expression trees based on the gene  $C = + / * aSb + bbababab$ . The decomposition of the first sub-expression tree is shown in Fig. 49.1, and more sub-expression trees by reading K-expression from the second head character beginning to the last are shown in Fig. 49.3.

As shown in Figs. 49.1 and 49.3, the gene C decomposed into five sub-expressions. The number of sub expressions is less than the head length of the gene, because the MGEP algorithm doesn't build an expression tree for a single node. Sub-expressions in Fig. 49.3 are not obtained only by separating the expression 49.1, it produced many novel expressions. The reason is that the character of the depth and breadth in the new sub expression trees are different. The tail length of the gene is larger than the required length, so the sub-expression is valid.



**Fig. 49.3** A K-expression translates into five sub-expressions. **a**  $E_2 = [b * \sin(b + b)]/a$ . **b**  $E_3 = a * \sin(b)$ . **c**  $E_4 = \sin(b)$ . **d**  $E_5 = b + b$

### 49.3.3 MGEP Fitness Assignment

The number of the sub-expressions of a gene is less than the head length of the gene. The MGEP gene fitness is defined by the best sub-expression. This fitness assignment idea is identical with the MEP algorithm. For instance, if we solve a symbolic regression problem, the fitness of each sub-expression  $E_i$  may be computed using the formula:

$$f(E_i) = \sum_{j=1}^N |O_{j,i} - w_j| \tag{49.2}$$

where  $O_{j,i}$  is the value of the sub-expression  $E_i$  of the gene on the  $j$ th sample data and  $w_j$  is the corresponding target result. The fitness of the gene  $f(G)$  is equal to the lowest fitness of the sub-expression in a single gene.

$$f(G) = \min f(E_i) \tag{49.3}$$

### 49.3.4 MGEP Genetic Operators

The MGEP technique uses the same search operators as the GEP, such as crossover, mutation and transposition to obtain new individuals. In crossover, two parent genes are randomly chosen and paired to exchange some material between them. Mutation can occur anywhere in the gene, but the structural organization of the gene should be preserved in this genetic operator. In MGEP, there are also three kinds of recombination: one-point, two-point and gene recombination. The genetic operation of MGEP is also operated on K-expression. So this novel technology does not damage the structure of the classical GEP.



### 49.3.5 MGEP Algorithm

In this section, we introduced the MGEP algorithm. The new algorithm only improved GEP decoding way and fitness assignment of a gene. A K-expression is decoded into several expression trees. So a gene of the MGEP contains multiple solutions for a problem. Combined with MEP idea the best solution is chosen for the MGEP gene. The MGEP algorithm is described in MATLAB language as follows:

```
function gene_fit = Fitness_Assignment(k_exp,T,h,t)
% Input arguments: a K-expression, terminal symbols, head length, tail length
% Output arguments: fitness value of the gene
% Create_ET is a another function which established an expression tree based on
% a sub-expression
% The Traverse function is to traverse an expression tree
% The Evaluate function is to assess a gene
for i = 1:h
    sub_k = []; sub_ET = []; sub_expression = []; %Initialize variables
    if(~isempty(findstr(k_expression(i),T))) % Death gene is not taken into account
        sub_k = k_expression(i:h+t); % Decomposition of a K-expression
        sub_ET = Create_ET(sub_k); % Create an sub-expression tree
        sub_expression = Traverse(sub_ET); % Traverse the sub expression tree
        sub_fit(i) = Evaluate(sub_expression); % Evaluate the sub expression
    end
end
gene_fit = min(sub_fit);
```

## 49.4 Experiments and Analysis

We conducted two experiments, and compared the MGEP evolutionary performance with the MEP and the GEP program. Three genetic algorithms were set with the same evolution parameters. The length of a MEP chromosome was set to be equal to the head length of a GEP gene. The number of MEP genes was more than the MGEP's. So we can compare their performance in the same environment. The main parameters are given in Table 49.1.

**Experiment:** Two symbol regression problems, as the following formula:

$$y = x^4 + x^3 + x \quad (49.4)$$

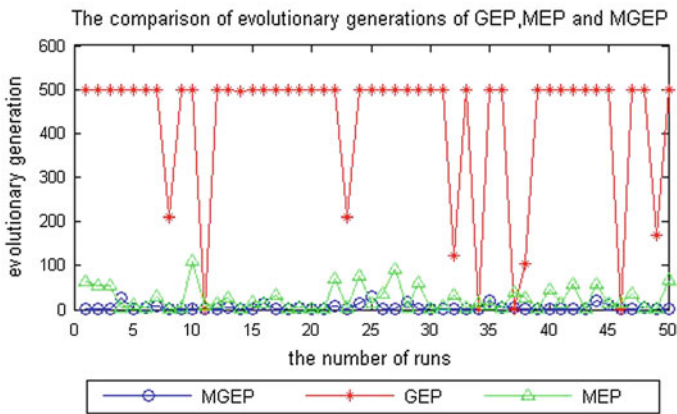
$$z = x^3 + x^2y + y \quad (49.5)$$

Based on the fourth problem, the terminal symbols would be set  $T = \{x\}$ , and the other function is set to  $T = \{x, y\}$ . In this two problems, we choose  $h = 10$ . For the sake of fairness, the length of a MEP chromosome takes the value 10. So

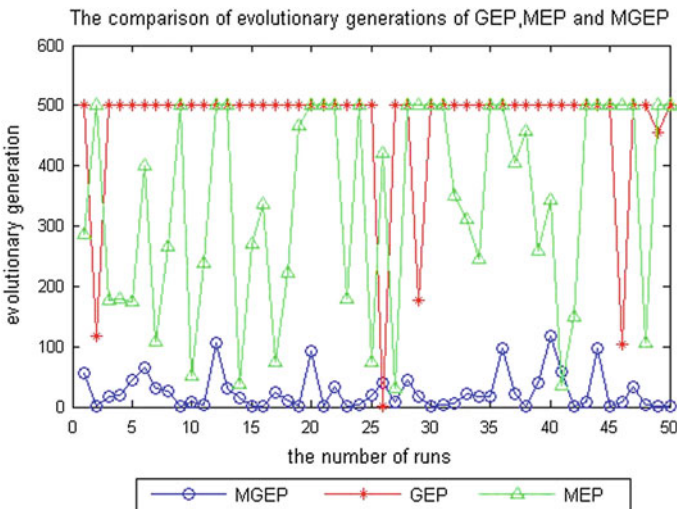
**Table 49.1** Two symbolic regression problems of experimental parameters

Function set	Number of chromosomes	Number of training data	Maximum generation	Crossover probability	Mutation probability	Accuracy
{+,*}	50	15	500	0.9	0.5	0.000001

MGEP and MEP algorithm includes approximately the same number of genes. Other parameters are given in Table 49.1. For two problems, each algorithm ran 50 times, and the comparison performance of three kinds of algorithm is shown in Figs. 49.4 and 49.5 respectively, the specific result in the following Tables 49.2 and 49.3.



**Fig. 49.4** The evolutionary generation of the expression  $y = x^4 + x^3 + x$



**Fig. 49.5** The evolutionary generation of the expression  $z = x^3 + x^2y + y$

**Table 49.2** Comparison evolutionary results of the three kinds of algorithm for the problem  $y = x^4 + x^3 + x$

Algorithm	Average generation	The longest generation	The shortest generation	Probability of success (%)
MGEP	4.68	28	1	100
Standard GEP	426.46	497	1	20
Standard MEP	25.64	110	1	100

**Table 49.3** Comparison evolutionary results of the three kinds of algorithm for the problem  $z = x^3 + x^2y + y$

Algorithm	Average generation	The longest generation	The shortest generation	Probability of success (%)
MGEP	25.46	117	1	100
Standard GEP	467.06	455	1	10
Standard MEP	343.12	466	31	58

The results of Experiment 1 showed that, in function mining with one variable, MGEP algorithm was obviously better than the standard GEP and MEP algorithm. The average number of generation to success of the MGEP is only 1.1 % of the GEP and 18 % of the MEP, and the success probability of function mining was significantly higher than the standard GEP and MEP algorithm.

In function mining with two arguments, the average number of generation to success of MGEP was about only 5.3 % of the GEP and 7 % of the MEP. These two experimental results showed the evolution efficiency of the MGEP algorithm was much higher than the standard GEP and MEP algorithm. Because the novel algorithm has the MEP unique feature which can store multiple solutions for a problem in a single gene.

## 49.5 Conclusion

This paper proposed a new genetic algorithm MGEP. And we introduced the fitness assignment and genetic operators for the MGEP. The algorithm integrates GEP and MEP idea which contains several sub-expressions. Experiments showed that when compared with the traditional GEP and the MEP, MGEP algorithm significantly improved the evolutionary performance.

**Acknowledgments** This work was supported by the National Natural Science Foundation of China (Grant No. 61170199), Hunan Provincial Innovation Foundation for Postgraduate (CX2012B367), Guangxi Key Laboratory of Trusted Software (Guilin University of Electronic Technology), and the Scientific Research Fund of Education Department of Hunan Province, China (Grant No. 11A004).

## References

1. Oltean M, Grosan C (2004) Evolving digital circuits using multi-expression programming. In: Proceedings of NASA/DoD conference on evolvable hardware. IEEE Press, pp 87–94
2. Oltean M, Grosan C, Diosan L, Mihaila C (2008) Genetic programming with linear representation a survey, WSPC/INSTRUCTION FILE
3. O’Neill M, Vanneschi L, Gustafson S, Banzhaf W (2010) Open issues in genetic programming. *Genet Program Evolvable Mach* 11:339–363
4. Oltean M, Grosan C (2003) A comparison of several linear genetic programming techniques. *Complex Syst* 14:285–313
5. He P, Kang L, Johnson CG, Ying S (2011) Hoare logic-based genetic programming. *Sci Ch Inf Sci* 54(3):623–637
6. Ferreira C (2002) Gene expression programming, 1st edn. Angra do Heroismo, Portugal
7. Ferreira C (2004) Gene expression programming and the evolution of computer programs. In: de Castro LN, Von Zuben FJ (eds) Recent developments in biologically inspired computing, vol 5. Idea Group Publishing, New York, pp 82–103
8. He P, Johnson CG, Wang HF (2011) Modeling grammatical evolution by automaton. *Sci Ch Inf Sci* 54(12):2544–2553
9. Karakasis VK, Stafylopatis A (2006) Data mining based on gene expression programming and clonal selection. IEEE congress on evolutionary computation, Vancouver, Canada, pp 514–521
10. Tsakonas A (2006) A comparison of classification accuracy of four genetic programming-evolved intelligent structures. *Inf Sci* 176:691–724
11. Cattani PT, Johnson CG (2010) ME-CGP: multi expression cartesian genetic programming. *IEEE Congr Evolut Comput* 2010:1–6
12. Yanan W, Bo Y, Zhao X (2009) Countour registration based on multi-expression programming and the improved ICP. *IEEE*, 2009
13. Chen Y, Jia G, Xiu L (2008) Design of flexible neural trees using multi expression programming. In: Proceedings of Chinese control and decision conference, vol 1, pp 1429–1434

# Chapter 50

## The Signal Processing Method of Mixed Interference Distributed Fiber-Optic Long-Distance Pipeline Leaks Detection System

Zhengsong Hu, Qihua Yang, Qiang Wang and Renjie Zhang

**Abstract** A new measuring structure based on the principle of Sagnac and Mach–Zehnder mixed distributed fiber-optic interferometer can detect pipeline leakage and locate leakage point in real-time. But the actual pipeline leakage signal is non-linear and tiny broadband signal, and there is a lot of background noise in the testing environment. The null spectrum extracted to decide the location of the leak is relatively difficult. In this paper, it's based on the discrete wavelet and least squares curve fitting method for leakage signal preprocessing. Reusing the signal through multi-scale decomposition of each signal multiplication method to determine leakage null spectrum, and the good result is achieved.

**Keywords** Distributed fiber-optic · Mixed interferometer · Leakage detection · Wavelet · Least squares curve fitting

### 50.1 Introduction

Pipeline transportation has the advantage of security, economy, non-polluting and convenient to manage. Currently, At home and abroad has been commonly adopted long-distance pipeline to transport the oil and gas. Distributed fiber sensing is a new technology developed in recent years, which has the advantage of long distance continuous monitoring, resistance to electromagnetic interference and long service life. The distributed fiber sensing also can replace the traditional single-point sensor array, then be used in long-distance pipeline detection systems [1]. Among it, zero frequency in the pipeline leakage detection signal determines the location of the leak point. But as pipeline leakage signal is non-linear, tiny

---

Z. Hu (✉) · Q. Yang · Q. Wang · R. Zhang  
College of Quality and Safety Engineering, China Jiliang University, Hangzhou,  
Zhejiang, China  
e-mail: woxindong0203@163.com

broadband signal, and there is a lot of background noise in the detection environment, which will result in null spectrum curve not smooth and bringing a lot of glitches. Then it will affect the position accuracy on the leakage point.

Taking advantage of traditional approach of non-stationary signals including Fourier transform, Wigner-Ville distribution and so on to analysis has certain limitations [2]. Especially Fourier transform to this multi-component signal will produce false signals due to cross-interference term, then affect the identification of leakage signal characteristics [3, 4].

Wavelet transform has a good time and frequency domain localization features and it is a new time–frequency analysis method, which is called as microscope in mathematical analysis. Yan Zhou analyzes principles and methods of wavelet thresholding denoising. Adopting the wavelet soft threshold denoising to process the detection signal in actual field conditions. Acquiring good noise cancellation, improving the detection sensitivity of the detection system effectively [5]. Jingjie Qian takes the advantage of discrete wavelet transform and linear least square curve fitting method to the frequency response of the Sagnac interferometer for pretreatment. Then the method which adopts threshold-derivation to determine the position of the notch point makes certain effect [6]. In order to achieve real-time monitoring of the pipeline leakage, this paper adopts a mixed interference distributed optical fiber measure structure of Mach–Zehnder and Sagnac. By taking advantage of discrete wavelet and least-squares curve fitting to do pretreatment for the leakage signal, then adopt leakage signal in the wavelet decomposition of the multiscale signal multiplication method to determine the null spectrum leakage signal, achieving good effect, providing a certain amount of a certain technical support for pipeline health monitoring.

## 50.2 Mixed Interference Detection System

The detection system adopts a Mach–Zehnder and Sagnac mixed fiber-optic measure structure, structure shown in Fig. 50.1. Including a light source (Laser), the optical circulator (CIR), the delay optical fiber, the sensing fiber, the optical coupler, the Faraday rotating mirror, the optical phase modulator and the photo-detector (PD). Compared with the the Dr. Shiju Huang sensing structure [7], The adding of the optical circulator (CIR) has two effects, on the one hand, to avoid the return light entering the laser, then case damage to the laser; on the other hand, to reduce the additional loss of the optical signal which is caused by traditional interference structure. Coupler DC1 adopts  $1 \times 2$  coupler to replace the original structure of the  $2 \times 2$  coupler, in order to reduce the losses of the way of the optical signal and to simplify the system structure [8, 9].

When the pipeline along the S at leak, The friction of fluid leakage and leakage hole wall will produce the pipe wall stress wave. The stress wave will produce disturbance to the sensing fiber which is laid on the wall, thereby affecting fiber

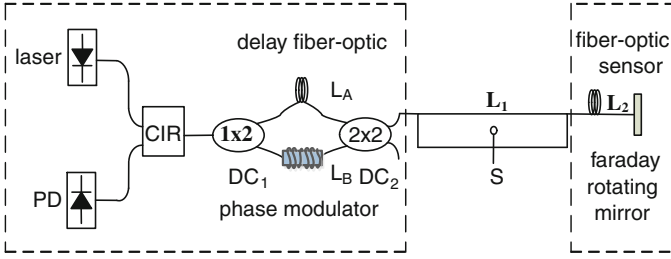


Fig. 50.1 Mixed distributed fiber-optic measure structure

diameter, length, Index of refraction, thus making an effect on the phase of the optical signal within the fiber [10, 11]. Its expression is as follows.

$$\Delta\phi = \Delta\phi_L + \Delta\phi_n + \Delta\phi_D \tag{50.1}$$

In this formula,  $\Delta\phi$  is phase Change of the lightwave,  $\Delta\phi_L$  is phase delay which is caused by the changes in the fiber length,  $\Delta\phi_n$  is the phase delay which is caused by the optical refractive index change.  $\Delta\phi_D$  is phase delay which is caused by the change in the optical fiber core diameter.

A phase modulator is installed in one arm as Fig. 50.1, modulating the high frequency carrier signal into the optical path, then can acquire the interference term expression of the light path one and path two [7].

$$I_0 = \cos\left\{4\phi_s \cos \omega_s \left(t - \frac{\tau_T}{2}\right) \sin \omega_s \left(\frac{\tau_A}{2}\right) \cos \omega_s(\tau_s) + \phi'_m \cos \omega_m t'\right\} \tag{50.2}$$

In this formula,  $\phi_s$  is the amplitude of the leakage points phase signal,  $\omega_s$  is the angular frequency of the leakage points,  $\tau_T$  is the time required for Light through the entire optical path,  $\tau_A$  is the time required for light passing through the delay fiber,  $\tau_s$  is the time required for light transmitting from the Leakage points to the Faraday rotator mirror,  $\phi'_m$  is the modulation signal amplitude,  $t' = t - \tau_T/2$ .

Adopt PGC demodulation circuit to demodulate, obtain an expression for the output of the interference signal is as follows.

$$\phi_X(t) = 4\phi_s \cos \omega_s \left(t - \frac{\tau_T}{2}\right) \sin \omega_s \left(\frac{\tau_A}{2}\right) \cos \omega_s(\tau_s) \tag{50.3}$$

In these three function entry,  $\cos \omega_s(t - \frac{\tau_T}{2})$  is the time domain signal, the other two is unrelated to time, do fast Fourier transform to (50.3), get the phase signal spectrum. In the leakage signal within the bandwidth range of 0–50 kHz [12], signal spectrum appears some null spectrum which contain the leakage location information. In the null spectrum,

$$\cos \omega_s(\tau_s) = 0 \tag{50.4}$$

It is found that

$$\omega_s(\tau_s) = \frac{(2k+1)\pi}{2} (k = 0, 1, 2, \dots) \quad (50.5)$$

Take  $k = 1$ , which takes the first null spectrum, it changes as follows.

$$L_s = \frac{c}{4 \cdot n \cdot f_s} \quad (50.6)$$

In this formula,  $c$  is the speed of light,  $n$  is the refractive index of the fiber core,  $f_s$  is null spectrum value,  $L_s$  is the distance between the leakage point and the Faraday rotation mirror. Location of the leakage point can be obtained by the value of the first null spectrum spectrogram, then to achieve the positioning of the leakage point.

### 50.3 Experiment and Analysis

Make FFT changes to demodulated interference signal to the system about  $4\phi_s \cos \omega_s(t - \frac{\tau_A}{2}) \sin \omega_s(\frac{\tau_A}{2}) \cos \omega_s(\tau_s)$ , Its unilateral spectrum is as the following formula [13].

$$\begin{aligned} F(\omega) &= 4\phi_s \sin \omega_s\left(\frac{\tau_A}{2}\right) \cos \omega_s(\tau_s) \delta(f - f_s) \\ &= 2 \left[ \sin \omega_s\left(\frac{\tau_A + 2\tau_s}{2}\right) + \sin \omega_s\left(\frac{\tau_A - 2\tau_s}{2}\right) \right] \end{aligned} \quad (50.7)$$

Known as the formula (50.7), Leakage signal spectrum is similar to the superposition of the two sinusoidal signals, among it,  $\sin \omega_s(\frac{\tau_A + 2\tau_s}{2}) + \sin \omega_s(\frac{\tau_A - 2\tau_s}{2})$  determine the trend of change in the signal.  $\tau_A$  is decided by the length of the delay fiber, for the system,  $\tau_A$  is constant value. When the pipeline leak,  $\tau_s$  is also constant value, such as the formula (50.7) as a time domain signal,  $\frac{\tau_A + 2\tau_s}{2}$ ,  $\frac{\tau_A - 2\tau_s}{2}$  are seen as two sinusoidal signal frequency,  $\omega_s$  is regarded as the time variable, known as the system architecture, the values of  $\tau_A$  and  $\tau_s$  are relatively small, then can see the values of  $\frac{\tau_A + 2\tau_s}{2}$  and  $\frac{\tau_A - 2\tau_s}{2}$  are also relatively small, so  $\sin \omega_s(\frac{\tau_A + 2\tau_s}{2}) + \sin \omega_s(\frac{\tau_A - 2\tau_s}{2})$  can be regarded as a superposition of the two low-frequency sinusoidal signal. Make analog pipeline leakage detection under laboratory conditions, the null spectrum diagram got as shown in following Fig. 50.2, as can be seen from the figure, the spectrum is equivalent to the low frequency slowly varying signal with the high frequency noise interference, conform to the theoretical analysis.

Due to the presence of noise interference, null spectrum curve has more glitches, not conducive to the identification, need to deal with noise, improve the identification of null spectrum. Because wavelet transform has the time domain and frequency domain localized, multi-resolution analysis, low entropy, very suitable for pipeline leakage signal analysis and processing.



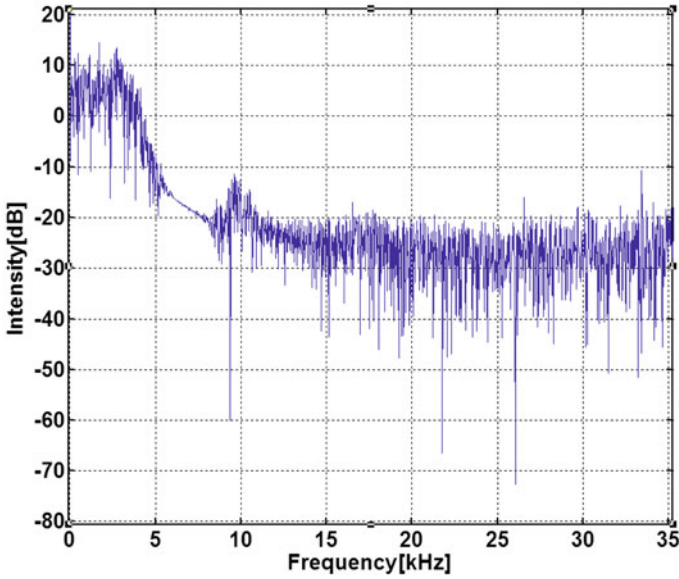


Fig. 50.2 Leakage signal null spectrum plot

Known as the wavelet transform principle [14], to any signal  $x(t) \in L^2(R)$ , its wavelet transform is defined as:

$$WT_x(a, b) = \langle x(t), \psi_{a,b}(t) \rangle = \frac{1}{\sqrt{a}} \int_R x(t) \psi^* \left( \frac{t-b}{a} \right) dt \quad (50.8)$$

In this formula,  $a$  is expansion factor,  $b$  is translation factor,  $\psi_{a,b}(t)$  is called Wavelet basis function depending on the parameters  $a$  and  $b$ ,  $\psi_{a,b}^*(t)$  is conjugate with  $\psi_{a,b}(t)$ , Symbol  $\langle \cdot \rangle$  represents Inner product.

In practical applications [15], wavelet transform need to be discretized, among continuous wavelet transform, the discretization formula of expansion factor  $a$  and translation factor  $b$  are taken for  $a = a_0^j$ ;  $b = ka_0^j b_0$ ,  $a_0$  and  $b_0$  are real constant which are greater than zero,  $j$  and  $k$  are integer.

Usually further to get constant as  $a_0 = 2$ ,  $b_0 = 1$ , the binary wavelet transform after discrete is as the following formula.

$$WT_x(j, k) = \langle x(t), \psi_{j,k}(t) \rangle = 2^{-j/2} \int_R x(t) \psi^*(2^{-j}t - k) dt \quad (50.9)$$

In order to eliminate the high-frequency noise, based on FFT transform, adopt wavelet transform to eliminate noise to the signal which is collected in the leakage sound field. Select sym5 wavelet, the signal is decomposed to the fifth layer, and choose minimax threshold for processing, reconstruct approximate coefficient and

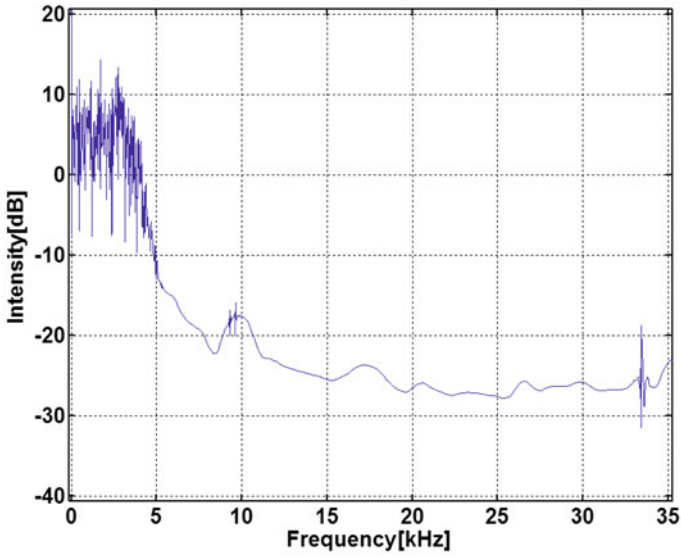


Fig. 50.3 Null spectrum after wavelet denoise

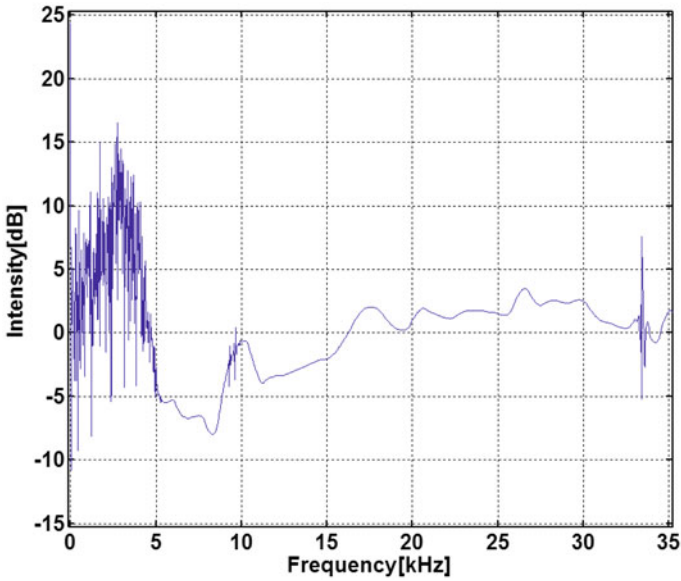


Fig. 50.4 Null spectrum after least squares fit

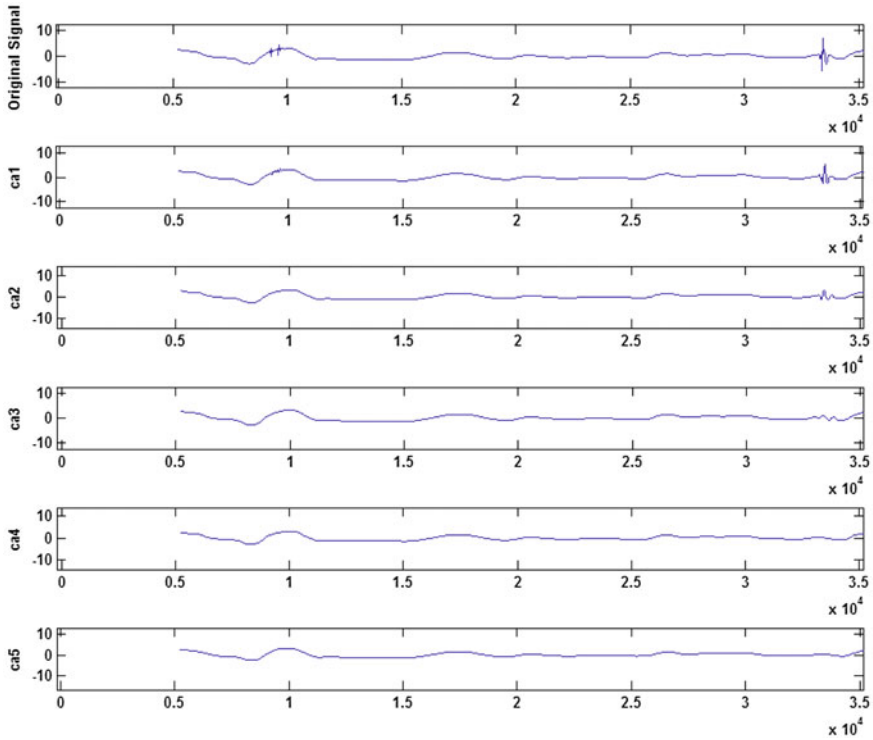
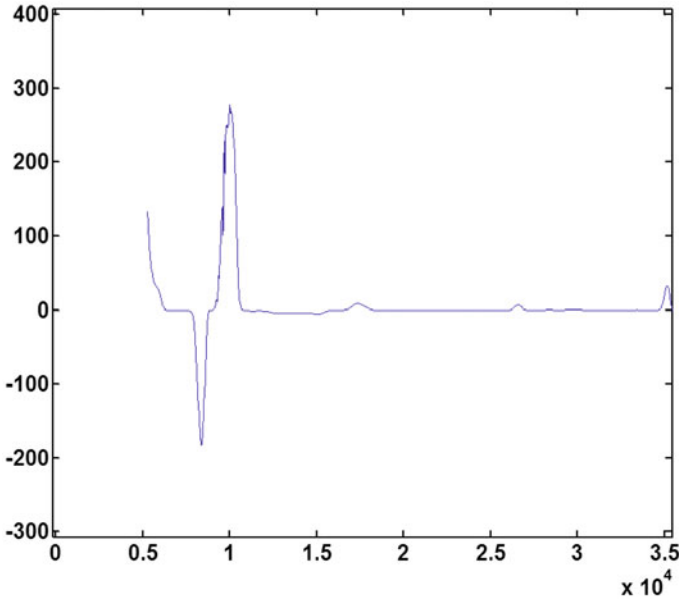


Fig. 50.5 Null spectrum signal wavelet decomposition approximation signal

get Fig. 50.3. Compared with Figs. 50.2 and 50.3, It can be seen, After wavelet consumer impatient, the change trend is very obvious, But there are still some noise, adopt one-dimensional blind Pathfinder optimization method may find Fig. 50.3 some glitches point, can not find the null spectrum accurately.

Therefore, it needs an effective, reliable signal processing method to accurately locate the leak point. This paper adopts discrete wavelet and least-squares curve fitting to do pretreatment for the leakage signal, then adopt leakage signal in the wavelet decomposition of the multiscale signal multiplication method to determine the null spectrum leakage signal, achieving good effect.

Before adopting the leakage in the wavelet decomposition of the multiscale signal multiplication method to determine the null spectrum, the leakage signal must be pretreated. Then leakage must form such as sin change, or aptly called shaped like caroline hill [6], as can be seen from Fig. 50.3, and compared with the desired signal, the collected leakage signal exists a significant downward trend, need to get rid of the downward trend, experiments show that least-squares curve fitting can remove the downward trend.



**Fig. 50.6** Null spectrum signal wavelet decomposition scale signal multiplied figure

Figure 50.4 is the least squares curve fitting null spectrum diagram, compared with Figs. 50.4 and 50.3, show that the least-squares curve fitting can be removed the leakage signal spectrum curve downward trend.

Can be seen from Fig. 50.4, after pretreatment, null spectrum signal containing the noise will still affect the finding of null spectrum. In front of a piece of data noise is in serious condition, in this paper, using the truncated data to deal with the null spectrum signal by sym5 multi-scale wavelet decomposition. From small-scale the  $ca_1$  into large scale  $ca_5$ , as shown in Fig. 50.5. After the decomposition, the effect of null spectrum signal curve is good, in addition to the noise, the curve is smooth, and it is more easier for null spectrum identification.

In order to further improve the identification accuracy of the null spectrum, using the null spectrum signal through multi-scale decomposition of each approximate signal multiplication [16], as shown in Fig. 50.6 ( $ca_1 * ca_2 * ca_3 * ca_4 * ca_5$ ) shows, the signal of null spectrum is fairly obvious, but there are several local extreme need further analysis confirmed. Therefore determined the current time in each scale wavelet transform signal multiplied, then compared with the previous time signal in each scale wavelet transform signal is multiplied by the value of the mean. If it exceeds a certain threshold, it will be judged as possible to null spectrum. The null spectrum is confirmed by further analyzing.

## 50.4 Conclusion

In this paper, it is based on the discrete wavelet and least squares curve fitting method for mixed distributed fiber-optic interferometer acquisition leakage signal preprocessing. Reusing the signal through multi-scale decomposition of each signal multiplication method to determine leakage null spectrum. Theoretical analysis and experimental data show that, this method can effectively for the identification of the null spectrum of fiber-optic interferometer signal which generated by pipeline leak. It provides a certain technical support for the realization of long distance pipeline leakage health monitoring.

**Acknowledgments** This project won the national natural science funds (60902095), National public welfare quality testing industry research project funding (201110058), National public welfare quality testing industry Special project funding (201310152), Zhejiang province natural science funds (Y1090672).

## References

1. Jiang Q, Wang Q (2012) Underwater gas pipeline leakage and location simulation in lab scale based on optical fiber interferometer. *J Cent South Univ* 43(2):576–580 (in Chinese)
2. Si Z (2011) Fourier transform and wavelet transform in the application of signal denoising. *Electron Des Eng* 19(4):155–157 (in Chinese)
3. Mark WD (1970) Spectral analysis of the convolution and filtering of non-stationary stochastic processes. *J Sound Vib* 11(1):19–63
4. Hu J, Zang L, Liang W, Wang C (2009) Based on the harmonic wavelet analysis of small pipe leakage diagnosis method. *J China Univ Petroleum* 33(4):118–124 (in Chinese)
5. Zhou Y (2006) Safety testing technology research of distributed optical fiber pipeline. Tianjin University, Tianjin, pp 67–81 (in Chinese)
6. Qian J, Zhang J, Jia B (2006) Real-time signal processing methods of Sagnac interference positioning system. *Chin J Sens Actuators* 19(4):1033–1037 (in Chinese)
7. Huang S, Lin W, Tsai M (2007) Fiber optic in-line distributed sensor for detection and localization of the pipeline leaks. *Sens Actuators A* 135(2):570–579
8. Qiao B, Yang Q, Wang Q (2011) Simulation analysis of improved interference distributed fiber-optic underwater gas pipeline leak detection. *J China Univ Metrol* 22(2):114–119 (in Chinese)
9. Hu Z, Yang Q, Qiao B (2012) The design of interference distributed fiber-optic underwater long gas pipeline leakage detection system. *Laser Optoelectron Prog* 49(070602):070602-1-070602-5 (in Chinese)
10. Tan J, Chen W, Wu J, Zhu Y (2008) Study on long distance pipeline destruction alarm technology based on Sagnac/Mach-Zehnder interferometers. *Acta Photonica Sinica* 37(1):67–72 (in Chinese)
11. Meng Ke (2004) Optical fiber interference measuring technology. Harbin Institute of Technology Press, Harbin, pp 44–45 (in Chinese)
12. Wassef WA, Bassim MN, Houssny-Emam M et al (1985) Acoustic emission spectra due to leaks from circular holes and rectangular slits. *Acoust Soc Am* 77(3):916–923 (in Chinese)
13. Zhao H, Hang L, Li G (2009) Pipeline leak positioning technology based on the optical fiber sensing and wavelet transform. *Transducer Microsyst Technol* 28(9):47–49 (in Chinese)

14. Liu S, Li Z, Wu J, Wang D, Liu J, Meng J (2008) Signal processing of displacement interferometer based on the continuous wavelet transform. *Chin J Lasers* 35(8):1235–1239 (in Chinese)
15. Hua W, Ma D, Wei J (2004) APD signal detecting method based on wavelet transform. *Chin J Lasers* 31(4):465–468 (in Chinese)
16. Dai X, Wu H (2002) Pendulum train tilting control system fault detection and diagnosis test research. *Electr Drive Locomot* 4:39–42 (in Chinese)

# Chapter 51

## Multiple Faces Tracking via Statistical Appearance Model

Jie Hou, Yaobin Mao and Jinsheng Sun

**Abstract** Recently, appearance based methods have become a dominating trend in tracking. For example, tracking-by-detection models a target with an appearance classifier that separates it from the surrounding background. Recent advances in multi-target tracking suggest learning an adaptive appearance affinity measurement for target association. In this paper, statistical appearance model (SAM), which characterizes facial appearance by its statistics, is developed as a novel multiple faces tracking method. A major advantage of SAM is that the statistics is a target-specific and scene-independent representation, which helps for further video annotation and behavior analysis. By sharing the statistical appearance models between different videos, we are able to improve tracking stability on quality-degraded videos.

**Keywords** Tracking · Appearance model · Learning

### 51.1 Introduction

Face tracking is a useful computer vision technology in the fields of human computer interface and video surveillance, and has been successfully applied in video conferencing, gaming and living avatars in web applications, etc. Facial

---

J. Hou · Y. Mao (✉) · J. Sun  
School of Automation, Nanjing University of Science and Technology,  
Nanjing, Jiangsu, China  
e-mail: myb\_nust@126.com

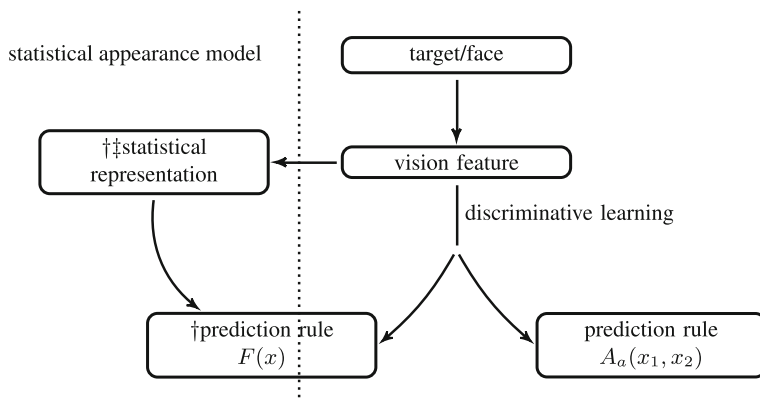
J. Hou  
e-mail: reiase@gmail.com

J. Sun  
e-mail: jssun67@yahoo.com.cn

appearance serves an important role in our daily communications, thus tracking human faces is usually the first step when a vision system tries to understand human beings.

Single face tracking has been well studied by researchers. Vacchetti [1] and Wang [2] use local interest-points matching to track 3D pose of human face. Their methods are robust against illumination and appearance variations, but suffer from error-prone feature matching. Zhou [3] proposed to track faces with Active Appearance Models (AAM), which has advantages of stability and alignment accuracy. One limitation of Zhou's method is that AAM fitting fails when handling large occlusions. Multiple faces problem is also common when performing video annotation and behavior analysis in real-world applications.

In this paper, we model facial appearance with the statistics (mean and variation) of the haar-like vision features, and train each face a discriminative prediction rule. An overview of statistical appearance model is shown in Fig. 51.1. The most significant difference between statistical appearance model and traditional methods is that we split the appearance model into two individual parts: statistical representation and prediction rule. The statistical representation is a target-specific and scene-independent representation. For discriminability, we learn a prediction rule that separating one face from the other faces with separability-maximum boosting (SMBoost). In our previous work [4], SMBoost shows better classification performance than AdaBoost and its online variation on UCI machine learning datasets, and achieves higher accuracy in tracking problems. The major contribution of our paper is a scene-independent appearance model. By reusing the statistical representations, we can apply the facial appearance model trained on one video sequence to other scenes.



**Fig. 51.1** † indicates that the model or the representation is target-specific, and ‡ indicates scene-independence. Prediction rule  $A_a(x_1, x_2)$  is used as global affinity measurement in multi-target tracking [5, 6], and  $F(x)$  is used to model posterior probability  $P(T|x)$  in tracking-by-detection [7]. Our work in this paper is described in the *left* part of the figure



## 51.2 Related Work

Tracking multiple targets that have complex motion traces is a challenging computer vision task, especially considering this problem in cluttered environments. Suppose we are tracking targets  $\{T_i\}_{i=1}^n$  in current frame, we can relocate a target  $T_i$  in the succeeding frame by searching for a candidate  $C$  from the scene that has maximum posterior probability  $P(T_i|C)$ . In tracking-by-detection, the tracker models  $P(T_i|C)$  with a binary classifier. Typically, haar-like feature and online boosting are used to train the classifier, for this combination has been proved to be very discriminative [8]. However, the classifier is scene-specified (see Table 51.1), for resampling of negative samples, which Viola used for generalizing [8], is not included.

In recent years, research on background subtraction and object detection have brought significant improvements to target detection, and lead to a trend of association based tracking, in which the target candidates  $\{C_j\}$  detected in succeeding frames are associated to the targets  $\{T_i\}$  in current frame. For simple scenes, the targets are represented with their motion models. By a precise estimating of motion states (e.g., position, speed and acceleration), we can guess the location of a target by likelihood probability  $P((x, y)|T_i)$ . With that guess, we can pair a candidate target  $C$  with the model that has maximum posterior probability  $P(T_i|C)$ .

However, motion model based association will be defeated if a target was absent from observation for a long time. In such situation, appearance is introduced as a secondary evidence of association. Some early approaches model the observations of a target directly with vision features of good invariance. Recently, Huang [9] proposes perform the association in a more adaptive manner. He joins tracklets into longer tracks by their linking probability:

$$P_{link}(T_i, T_j) = A_m(T_i, T_j) A_a(T_i, T_j) A_t(T_i, T_j) \quad (51.1)$$

where  $T_i$  is the target of the  $i$ th tracklet, and  $A_m$ ,  $A_a$ ,  $A_t$  indicate the motion, appearance and temporal affinities between the tracklets. The linking probability  $A_a(T_i, T_j)$  is a global appearance affinity measurement. A significant advantage of Huang's method is that the appearance affinity could be modeled with discriminative learning. Cheng [5] suggests to represent target appearance with local image descriptors, and use AdaBoost for feature selection. Low-supervised learning, for

**Table 51.1** A comparison of appearance models

		Target-specific	Scene-independent
Tracking-by-detection	Boosting [7]	+	-
	MIL [11]	+	-
Multi-target tracking	Global affinity measurement [5, 6]	-	-
	PIRMPT [12]	+	-
Our method		+	Partly

example, multi-instance boosting, is introduced to handle noisy data [10]. Table 51.1 is a summary of appearance models mentioned in this paper. We focus on whether the method models  $P(T_i|C)$  directly (target-specific), and scene-independence.

### 51.3 Statistical Appearance Model

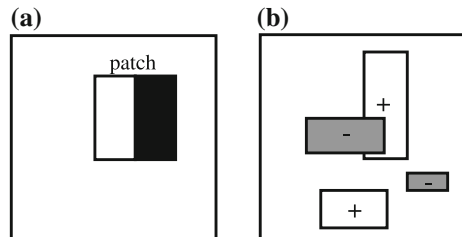
In this section, we present a statistical approach to model object appearance. The key idea of statistical appearance model is characterizing an object with the statistics of vision features (statistical representation, SR). Statistics is not a representation of good invariance. However, we are able to learn a discriminative prediction rule (PR) from the statistics if the vision feature (VF) is choosing properly and learning algorithm (LA) is carefully designed. Vision feature, statistical representation and learning algorithm are three important aspect of statistical appearance model. We will introduce each of them in detail.

#### 51.3.1 Vision Feature

Various vision features has been discussed for modeling appearance, e.g., shape, color histogram and texture (HOG) [12]. However, these features focus on invariance of a target, and might not discriminative enough to classify very similar targets (faces in this paper). Thus, researchers propose extracting these features on pre-defined regions to make the feature more discriminative [5].

Another method of enhancing discriminability is exploring an over-rich weak feature set with boosting. By combining weak classifiers of haar-like features, Viola builds the first practical real-time face detector. Original haar-like features that Viola used in his work are designed to capture within-patch structure patterns (see Fig. 51.2). Babenko uses non-adjacent haar-like feature, which combines 2–4 randomly generated rectangle regions and ignore the adjacency, in their work [10].

**Fig. 51.2** Haar-like feature.  
**a** Within-patch haar-like feature. **b** Cross-patch haar-like feature



### 51.3.2 Statistical Representation of Appearance

In this paper, we use statistical representation for characterizing facial appearance. The statistical representation of a target includes the mean  $E$  and variation  $\Sigma$  of the samples.

Suppose  $\{x_i\}$  is a data sequence, and the expectation of the first  $n$  samples is denoted by  $E[x]^{(n)}$ . When there comes a new sample  $x_{n+1}$ , we update the expectation  $E[x]$  with a learning rate  $\gamma$ .

$$E[x]^{(n+1)} = (1 - \gamma)E[x]^{(n)} + \gamma x \quad (51.2)$$

Variance  $\sigma^2[x]$  can be updated by a subtraction of two expectations:

$$\begin{aligned} \sigma^2[x]^{(n)} &= E[x^2]^{(n)} - (E[x]^{(n)})^2 \\ \sigma^2[x]^{(n+1)} &= (1 - \gamma)E[x^2]^{(n)} + \gamma x^2 - (E[x]^{(n+1)})^2. \end{aligned} \quad (51.3)$$

### 51.3.3 Learning Algorithm

In appearance base tracking, on-line boosting is a common choice for learning the prediction rules. Boosting chooses important features from the given feature pool, so that the prediction rule remains simple as it covers as many features as possible. However, the criterion function of AdaBoost (51.4) is difficult to be estimated on-line, for both the sample set  $\{(x, y)\}$  and the decision function  $F$  change in on-line learning paradigm [4].

$$F^* = \arg \min_F E_{\{(x,y)\}} [L(y, F(x))] \quad (51.4)$$

In our previous work [4], a new online boosting using separability based loss function instead of margin based loss function, is designed. The separability based loss function characterizes the degree that the decision function  $F$  separates the samples of class  $c(\{(x, y)\}_{y=c})$  from the rest  $(\{(x, y)\}_{y=\bar{c}})$ .

$$\hat{L}(c, L, E[x]|_{y=c}, \Sigma[x]|_{y=c}), \quad c \in \{+1, -1\} \quad (51.5)$$

Separability-maximum boosting (SMBBoost) maximizes separability of both classes.

$$\begin{aligned} F^* &= \arg \min_F \Phi(F) \\ &= \arg \min_F E_c [L(c, F, E[x]|_{y=c}, \Sigma[x]|_{y=c})] \end{aligned} \quad (51.6)$$

$E[x]$  and  $\Sigma[x]$  in (51.6) denote the mean and variation of the samples, which are well estimated by (51.2) and (51.3). Within on-line learning paradigm, it is much

easier to optimize SMBoost (51.6), for both  $E[x]$  and  $\Sigma[x]$  are fixed when the algorithm searching for optimal  $F$ .

### 51.4 Multiple Faces Tracking

Faces are similar to each other, thus finding facial appearance differences is much more difficult than separating one face from its surrounding background. The fundamental problem of multiple faces tracking is separating the faces from each other. We build an association based multiple faces tracker with statistical appearance model in the paper. The overview of our tracking framework is shown in Fig. 51.3. Our scheme is similar with [12] except the appearance model. In our tracker, we do not use tracklets for simplicity and real-time performance. We associate face candidates in succeeding frame to the faces in current frame directly. Each step in our framework is described below.

**Face detection** Face detection takes a new frame from the test video sequence and applies a state-of-art face detector to it. Face detection produces face candidates  $\{C_j\}_{j=1}^c$  under association.

**Face classification** Denote by  $\{T_i\}_{i=1}^t$  the statistical representation of the faces already known, also denote by  $\{F_i(x) \in [-1, +1]\}_{i=1}^t$  the prediction rules, where

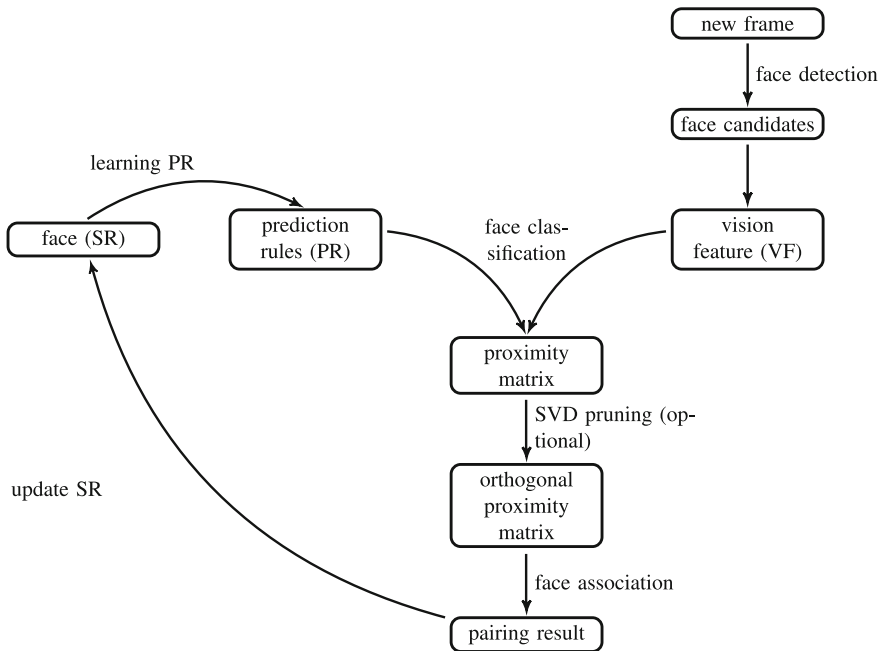


Fig. 51.3 System overview

$F_i(x)$  separate face  $T_i$  from the rest faces. We obtain the proximity matrix  $P$  by applying the prediction rules to the face candidates:

$$P_{i,j} = P(T_i|C_j) = \begin{cases} F_i(C_j), & F_i(C_j) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (51.7)$$

**SVD pruning (optional)** For face association, we need one-against-one pairing, which requires a successful pairing satisfy

$$P_{i,j} = \arg \max_{\delta} P_{\delta,j} = \arg \max_{\delta} P_{i,\delta}$$

Thus we use the method mentioned in [13] to prune the proximity matrix  $P$  into an orthogonal matrix.

**Face association** After pruning, we can associate the face candidates to the faces by maximizing posterior probability  $P(T_i|C_j) = \widehat{P}_{i,j}$ .

**Update statistical representation** For faces that have successfully paired with candidates in succeeding frame, we update their statistical representation with a learning rate  $\gamma$ .

**Learning prediction rule** Suppose  $\{T_i\}_{i=1}^t$  are the faces that we are tracking, we learn each face a discriminative prediction rule. When training the rule, we use statistical representation of  $T_i$  as positive SR, and combine the statistical representations of the rest faces as negative SR  $\overline{T}_i = \sum_{\sigma \neq i} T_{\sigma}$ .

## 51.5 Experiments

In this section, we first perform experiments demonstrating the effectiveness of SAM for multiple faces tracking problems. Then, we present an experiment of sharing statistical representation between two video sequences.

### 51.5.1 Tracking Multiple Faces

We first experiment our method for multiple faces tracking problems. Two test sequences from [14] are used in our evaluation. Association fails when handling face candidates with scaling factor larger than two. Thus we stop the association when one of the detected faces is too large. The tracker will re-catch the faces after scaling. Since we use a motion-free target model, our tracker is robust against large occlusions and missing detecting, and tracks the faces stably. The tracker may assign candidates to wrong faces when facial appearance varies too much, e.g., laughing and rotating. In such situation, we should stop updating the statistical appearance representation. And the faces will switch back when the facial



**Fig. 51.4** Tracking multiple faces on Seq. face\_fast [14]



**Fig. 51.5** Tracking result on Seq. face\_frontal

appearance stops varying. Such failures could be reduced by rejecting sudden large moves (Figs. 51.4, 51.5).

### 51.5.2 Sharing Statistical Representations Between Video Sequences

In this experiment, we share the statistical representations when tracking human faces on different video sequences. Sharing appearance model is useful for video annotation and behavior analysis.

Two test video sequences are used in this experiment. Seq. nokia is a quality-degraded video sequence captured with a Nokia smart phone, which suffers shaking and motion blur (intentionally). Seq. samsung is captured with another Samsung smart phone. Fig. 51.6 shows tracking result of our method on Seq. nokia. And Fig. 51.7 presents tracking result of sharing the models. In Seq. nokia, our tracker fails to assign the candidates to the faces on frame 500 (associating fails), and miss one face in frame 140 and frame 300. By sharing the statistical representation estimated on a clearer observation (Seq. samsung), the stability of tracking result on Seq. nokia got improved, and find correct assignments on the failure frames.



**Fig. 51.6** Tracking multiple faces on Seq. nokia



**Fig. 51.7** Tracking result of sharing statistical representation. The *above* row shows tracking result on Seq. nokia, and the *row below* shows tracking result on Seq. samsung

## 51.6 Conclusion

In this paper, a statistical appearance model, which characterizes facial appearance with statistics, is proposed. SAM captures appearance invariance by exploring an over-rich haar-like feature set, and trains a classifier of good discriminability using separability-maximum booting. A novel framework using SAM is designed for multiple faces tracking. In our framework, we are able to track faces robustly. By sharing the statistical representations, we are able to improve tracking result on quality-degraded video sequence.

**Acknowledgments** This work is supported by the National Natural Science Foundation of China (Grant No. 60974129) and Intramural Research foundation of NJUST (2011YBXM119).

## References

1. Vacchetti L, Lepetit V, Fua P (2004) Stable real-time 3D tracking using online and offline information. *IEEE Trans Pattern Anal Mach Intell* 26:1385–1391
2. Wang Q, Zhang W, Tang X, Shum H-Y (2006) Real-time bayesian 3-D pose tracking. *IEEE Trans Circuits Syst Video Technol* 16:1533–1541
3. Zhou M, Liang L, Sun J, Wang Y (2010) AAM based face tracking with temporal matching and face segmentation. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, 2010, pp 701–708
4. Hou J, Mao Y, Sun J (2012) Visual tracking by separability-maximum online boosting. In: *12th International conference on control automation robotics vision (ICARCV)*, 2012, pp 1053–1058
5. Kuo C-H, Huang C, Nevatia R (2010) Multi-target tracking by online learned discriminative appearance models. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, 2010, pp 685–692
6. Yang B, Nevatia R (2012) Multi-target tracking by online learning of non-linear motion patterns and robust appearance models. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, 2012, pp 1918–1925
7. Grabner H, Grabner M, Bischof H (2006) Real-time tracking via online boosting. In: *Proceedings of BMVC*, pp 47–56
8. Viola P, Jones M (2002) Robust real-time object detection. *Int J Comput Vis* 57:137–154

9. Huang C, Wu B, Nevatia R (2008) Robust object tracking by hierarchical association of detection responses. In: Forsyth D, Torr P, Zisserman A (eds) *Computer vision—ECCV*. Springer, Berlin, pp 788–801
10. Babenko B, Belongie M-HYS (2011) Robust object tracking with online multiple instance learning. In: *IEEE transactions on pattern analysis and machine intelligence (TPAMI)*
11. Babenko B, Yang M, Belongie S (2009) Visual Tracking with Online Multiple Instance Learning. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL
12. Kuo C-H, Nevatia R (2011) How does person identity recognition help multi-person tracking? *IEEE conference on computer vision and pattern recognition (CVPR)*, pp 1217–1224
13. Delponte E, Isgrò F, Odone F, Verri A (2006) SVD-matching using SIFT features. *Gr Models* 68:415–431
14. Maggio E, Piccardo E, Regazzoni C, Cavallaro A (2007) Particle PHD filtering for multi-target visual tracking. Presented at the *IEEE international conference on acoustics, speech and signal processing*



# Chapter 52

## A Fast Dictionary Training Algorithm for Single Image Super-Resolution

Yiliang Lv and Jiwei Liu

**Abstract** Generally the dictionary for single image super-resolution is trained by iterations of MP algorithm and K-SVD algorithm. Using the dictionary, low resolution images can be restored to high resolution images with high quality. But the training process always takes a lot of time. So in this paper we use SVD to analyze the space relationship between the high and low resolution samples, and present a cluster based algorithm for dictionary training. Compared with the K-SVD based algorithm, the proposed algorithm trains the dictionary with a much higher speed, and restores the images with similar visual quality.

**Keywords** Image super-resolution · Cluster · SVD · Spares representation

### 52.1 Introduction

Super resolution image reconstruction technique is really active these years. It helps to improve the resolution of digital image through the software, breaking the resolution limitation of the low-cost sensors, such as surveillance camera, synthesis aperture radar etc. It has large potential in the field of medical image diagnosis, identity recognition and satellites image analysis, the restored high resolution images which contain the low resolution missing details help people to do better decision.

The main idea of the super resolution reconstruction is to add prior knowledge to low resolution images, generally there are two ways. One is to add the prior knowledge by multi-sample. In this way some low resolution images which are generated with different Point Scale Function (PSF) or shot from different part are

---

Y. Lv (✉) · J. Liu

School of Automation and Electrical Engineering, University of Science and Technology Beijing, The 92th mailbox, No 30 XueYuan Road, Haidian District, Beijing, China  
e-mail: lv19890619@163.com

needed. To restore the image, firstly low resolution images are restored to high resolution images one by one by motion estimation and inverse PSF. Then transform the restored images to constraint conditions, which are used to generate the final high resolution image by MAP [1] or POCS [2] algorithms. But the algorithm has some drawbacks: in practice the conditions of low resolution images can't be fulfilled, and the motion estimation can't be calculated accurately, the worst is that the number of low resolution images is enormous and the image can't be restored with a high quality, when the magnification is large. So scientists found another way, to restore image through learning. In this way high resolution image is restored from single low resolution image and the prior knowledge extracted by learning from the preset samples. The prior knowledge can be stored into several data structures. At first Simon Baker and Takeo Kanade present hallucination algorithm which use pyramid structure to store the prior knowledge to restore human faces [3]. Later Yang presents to restore the image through the raw image patches [4] and trained dictionary [5] with the support of compressed sensing and sparse representation [6, 7]. Compared with the former one, learning based super resolution algorithm looses the condition and keeps the quality when the magnification is large.

In the learning based super resolution reconstruction, dictionaries are trained by iterations of MP algorithm [8] and K-SVD algorithm [9, 10]. The algorithm is reliable, and the quality of restored image is satisfying, but the computational complexity is high and the algorithm produce hallucination easily because of the over learning. In this paper we present a cluster based dictionary training algorithm which restores image with high quality, trains the dictionary with less computational complexity and resists to the hallucination phenomenon.

The remainder of the paper is organized as follows: Sect. 52.2 analysis the space relationship when down sampling and restoring, Sect. 52.3 presents a new method to train the super-resolution dictionary, Sect. 52.4 validates the effectiveness of the presented algorithm and in Sect. 52.5 we draw the conclusion.

## 52.2 Space Analysis

Let  $X_l = \{x_{l1}, x_{l2}, \dots, x_{lm}\}$  denotes the patches from the low resolution samples and  $X_h = \{x_{h1}, x_{h2}, \dots, x_{hm}\}$  denotes the corresponding patches from the high resolution samples. According to the image-forming principle, they shall follow the relationship.

$$X_l = D_s \cdot M \cdot B \cdot X_h + E \quad (52.1)$$

where  $D_s$ ,  $M$ ,  $B$ ,  $E$  stand for down sampling matrix, local movement matrix, blurring matrix and noises respectively. If we make the assumption that all the patches share the same blurring function and the movement is known. According to [5, 8, 11], if we deal the restoration problem via spares representation, the anti-

noise properties of the algorithm helps to keep noise away. So the Eq. (52.1) can be simplified as:

$$X_l = A \cdot X_h \quad (52.2)$$

where  $A$  is a constant matrix with  $\dim(x_{li})$  rows and  $\dim(x_{hi})$  columns. The restoration problem can be seen as an equation, in the equation  $X_h$  is the unknown,  $X_l$  is the known and  $A$  is the parameters of  $X_h$ . If we solve Eq. (52.2) with  $\dim(x_{hi}) > \dim(x_{li})$ , the solution should consist of a deterministic part and its kernel part. In order to estimate the kernel, we have to separate the two part of the solution. Do an SVD on (52.2). The equation can be turned into:

$$U_A^T \cdot X_l = [\Lambda_A \quad 0] \cdot V_A^T \cdot X_h \quad (52.3)$$

where matrixes  $U_A, V_A$  are orthogonal bases from high and low resolution patches space respectively. Matrix  $[\Lambda_A \quad 0]$  is a diagonal matrix, it helps to separates the deterministic part and the kernel part.  $\Lambda_A$  is singular, because every blur function for every pixel from the low resolution image is independent to each other in the spatial domain. According to [6–8], high resolution patches and low resolution patches share the same sparse representation, then let  $D_l, D_h$  represents the low and high resolution dictionary and  $S$  represents the corresponding sparse representation of  $X_l$ , they follow the relationship:

$$\begin{cases} X_l = D_l \cdot S \\ X_h = D_h \cdot S \end{cases} \quad (52.4)$$

Then put (52.4) in (52.3), so

$$D \cdot S = U_A^T \cdot X_l = [\Lambda_A \quad 0] \cdot V_A^T \cdot X_h = [\Lambda_A \quad 0] \cdot V_A^T \cdot D_h \cdot S \quad (52.5)$$

where  $D = U_A^T \cdot D_l$ , and then multiple  $\Lambda_A^{-1}$  on each side we get

$$[I \quad 0] \cdot V_A^T \cdot D_h \cdot S = \Lambda_A^{-1} \cdot D \cdot S = [I \quad 0] \begin{bmatrix} \Lambda_A^{-1} \cdot D \\ \tilde{D} \end{bmatrix} \cdot S \quad (52.6)$$

So high resolution dictionary  $D_h$ , can be written as  $V_A \cdot \begin{bmatrix} \Lambda_A^{-1} \cdot D \\ \tilde{D} \end{bmatrix}$ .

If the matrix  $A$  is known, the high resolution dictionary can be composed of two parts, one is the deterministic part from the low resolution dictionary, the other is from the kernel ( $A$ ). So with the form (52.6), we can find a faster way to get the dictionary.

### 52.3 Dictionary Learning

In sparse representation based super resolution algorithm, we need two redundant dictionaries, one for low resolution patches and the other for the high ones. Each column of the dictionary called an atom. And the atoms which share the same column number in each dictionary are corresponding. In order to search the atoms, generally we solve the problem:

$$\min_D \|S\|_0 \text{ subject to } \|X - D \cdot S\|_2 \leq \theta \quad (52.7)$$

where  $S$  stand for the sparse representation found by MP algorithm. From another part of view, atoms from  $D$  can be regarded as the redundant bases in the space, and  $S$  is the coordinate of the bases. The problem of searching the dictionary atoms can be translated to the problem of finding redundant bases.

Clustering on the hyper sphere is one way to find redundant bases, and it won't need much computation resources compared with the K-SVD. The sparsity of the representation is ensured because the bases are redundant, and the residual  $R = X - D \cdot S$  can be limited if the bases are complete. According to the compressed sensing theory, the restored high resolution image can be expected to have a high quality. Meanwhile MP algorithm also match with the cluster training dictionary, because if the sparse representation is found by the MP algorithm through the cluster training dictionary, there should be one dominant nonzero sparse representation coefficient, the other should be much smaller, intuitively it helps to keep the main part of high frequency details, prevent the mixture of them.

In order to train the dictionary, the samples are preprocessed, the low resolution samples and high resolution samples are transform to  $\Lambda_A^{-1} \cdot U_A^T \cdot X_l$  and  $V_A^T \cdot X_h$ , so that they can share the same part according to the Eq. (52.6). In order to keep the deterministic part of the high resolution image precise during the restoration, the low resolution dictionary shall be constructed first. So we have to cluster on the normalized low resolution samples and use the center of the classes to form the dictionary. Then weighted average the kernel part of the samples which are clustered in one class to form the kernel part of the high resolution dictionary. The weight coefficients are determined by  $w_{ij} = \frac{p_{ij}}{\sum p_{kj}}$ , where  $p_k$  is the  $k$ th low resolution sample's projection on the  $j$ th class's center.

Different from the K-SVD algorithm searching the atoms in the whole space, the cluster based algorithm search the atoms in the sample space, so that it keeps the atoms among the samples, and the restored image in the sample space, but the K-SVD algorithm won't, and that is why the hallucination come out less, if the cluster based algorithm is used.

Considering of image visual precision, in practical use, we use the K-MEANS cluster to choose the atoms, and use the dot product of uniformed samples to measure the distance.

The procedures for dictionary training are shown as follows:

Input: High resolution samples  $X_h$ , low resolution samples  $X_l$ , the degeneration matrix  $A$  and the dictionary size  $N$ .

Step 1: Do SVD on  $A$  and get  $V_A$ ,  $U_A$  and  $\Lambda_A$ .

Step 2: Transform the samples to  $\tilde{X}_l = \Lambda_A^{-1} \cdot U_A^T \cdot X_l$ ,  $\begin{bmatrix} \hat{X}_h \\ K \end{bmatrix} = V_A^T \cdot X_h$ ,  $\hat{X}_h$

and  $\tilde{X}_l$  are almost the same.

Step 3: Use K-MEANS cluster  $\tilde{X}_l$  to  $N$  class and use the center of class to form the dictionary  $D$ .

Step 4: Calculate  $\tilde{D}$  with  $\tilde{D} = K \cdot W$  where

$$w_{ij} = \begin{cases} \frac{p_{ij}}{\sum_{x_{ki} \text{ belong to the } j\text{th class}} p_{kj}} & \text{if } x_{li} \text{ belong to the } j\text{th class} \\ 0 & \text{else} \end{cases}.$$

Output: Low resolution dictionary  $D_l = U_A \cdot D$ , and high resolution dictionary  $D_h = V_A \cdot \begin{bmatrix} \Lambda_A^{-1} \cdot D \\ \tilde{D} \end{bmatrix}$ .

And the procedures for restoration are shown as follows:

Input: Low resolution patches  $X_l$  dictionaries  $D_h, D_l$ .

Step 1: Use MP algorithm to find sparse representation  $S$  with input  $X_l$  and  $D_l$ .

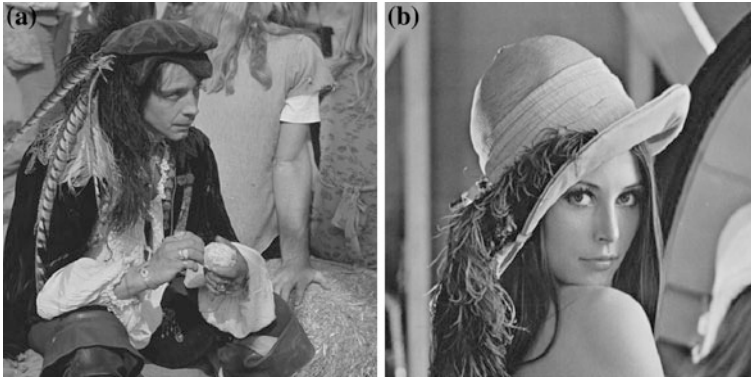
Step 2: Calculate  $X_h$  with  $X_h = D_h \cdot S$ .

Output: High resolution patches  $X_h$ .

## 52.4 Experiment Results

In order to validate the effectiveness of the cluster based dictionary training algorithm, we compare it with the K-SVD based dictionary training algorithm in two aspects one is the restored image quality, the other is the training time, use the PSNR and human visual sense to assess the quality of the restored image and training time to measure the training speed of the algorithm.

In the experiment, we use the patches sampled from the Fig. 52.1a to train the dictionary, and use the dictionary to restore the low resolution version of Fig. 52.1b. The size of the high resolution image is  $512 \times 512$ , the magnificent is 4, the size of



**Fig. 52.1** The high resolution images for experiment, **a** is used for sampling and **b** is for restoration



**Fig. 52.2** The result of the images: **a** original high resolution image, **b** cluster based algorithm's result, **c** K-SVD based algorithm's result

low resolution patches, the number of dictionary atoms, and sparsity of the representation are set to be  $3 \times 3$ , 1000, and 5 respectively.

In restoration there is one overlapped column or row between the adjacent low resolution patches, and all the sparse representation is found by MP algorithm. The iteration times of the two algorithms are 40 for the K-SVD based algorithm and 100 for cluster based algorithm.

We use Matlab to program the algorithm, and run it on a laptop with Intel(R) Core(TM) i5 CPU M480.

To start with the experiment, blur the high resolution image with a specific Gaussian function, and down sample the images to  $256 \times 256$ , then randomly select 22500 corresponding patch pairs from the Fig. 52.1a, and down sampled one. Initial a dictionary with raw samples for K-SVD based algorithm and using the samples to train the dictionary with the two algorithms separately. At last restore Fig. 52.1b with two dictionaries the result is shown in Fig. 52.2, and corresponding statics are in Table 52.1.

**Table 52.1** The table of training time, PNSR of restored image Fig. 52.1a, PNSR of restored image Fig. 52.1b, iteration times statics

Heading	Dictionary training time (s)	PNSR of restored image Fig. 52.1a	PNSR of restored image Fig. 52.1b	Iteration times
Cluster based algorithm	887.4	30.7363	33.1316	100
K-SVD algorithm	11858.1	30.7292	32.7347	40

Though the PSNR of the cluster based algorithms are little higher, from the visual sense, the K-SVD algorithm's result is sharper, but there is some hallucination near the contour of the hat and human body. And the cluster based algorithm reduced much dictionary training times compared with the K-SVD based algorithm. In all, the quality of image that restored by our algorithm is acceptable. And the computational complexity is much lower.

## 52.5 Conclusion

In this paper we present a method to train the dictionary for single image super-resolution reconstruction. Compared with the dictionary trained by K-SVD algorithm, the dictionary trained by our method restores images with similar quality, and our algorithm has a much higher speed for dictionary training. Because the training algorithm is based on clustering algorithm, we can determine the optimal dictionary size by Competitive Agglomeration algorithm, and modify the dictionary size easily with merging and dividing. These are what K-SVD algorithm can't achieve. However none of the analysis suggests how to find the optimal sparsity for the sparse representation of the patches. In further investigation we want to join manifold learning and image super resolution together in order to increase the image quality and restoration speed.

## References

1. Belekos S, Galatsanos N, Katsaggelos A (2010) Maximum a posteriori video super-resolution using a new multichannel image prior. *IEEE Trans Image Process* 19(6):1451–1464
2. Fan C, Zhu J, Gong J, Kuang C (2006) POCS super-resolution sequence image reconstruction based on improvement approach of Keren registration method. In: *Proceedings of 6th international conference ISDA*, 16–18 Oct 2006, pp 333–337
3. Baker S, Kanade T (2002) Limits on super-resolution and how to break them. *IEEE Trans Pattern Anal Mach Intell*, pp 1167–1183
4. Yang J, Wright J, Huang T, Ma Y (2008) Image super-resolution as sparse representation of raw image patches. In: *Proceedings of IEEE conference computer vision and pattern recognition*, pp 1–8
5. Yang J, Wright J, Huang J, Huang T (2010) Image super resolution via sparse representation. *IEEE Trans Image Process* 11(19):2861–2873

6. Donoho DL (2006) Compressed sensing. *IEEE Trans Inf Theory* 52(4):1289–1306
7. Rauhut H, Schnass K, Vandergheynst P (2008) Compressed sensing and redundant dictionaries. *IEEE Trans Inf Theory* 54(5):2210–2219
8. Moghadam AE, Shirani S (2007) Matching pursuit-based region-of-interest image coding. *IEEE Trans Image Process* 16(2):406–415
9. Aharon M, Elad M, Bruckstein A (2006) K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Process* 54(11):4311–4322
10. Liu J, Wu D (2011) Joint POCS method with compressive sensing theory for super-resolution image reconstruction. In: 3rd international conference on awareness science and technology, iCAST 2011, pp 99–102
11. Dong W, Li X, Zhang L, Shi G (2011) Sparsity-based image denoising via dictionary learning and structural clustering. In: IEEE conference computer vision and pattern recognition, pp 457–464



# Chapter 53

## Inner-Knuckle-Print Verification Based on Guided Image Filtering

Ming Liu and Jun Yan

**Abstract** This paper presents a new approach for inner-knuckle-print verification. Firstly, guided image filtering is implemented to remove noise and the minute lines. Then robust line features are extracted from the image based on a derivative edge detector. Finally the binary line images are matched by using a cross-correlation-based method. The experiments on a finger image database which includes 2000 images from 100 different individuals show good performance of the proposed approach.

**Keywords** Biometrics · Inner-knuckle-print · Guided image filtering · Cross-correlation

### 53.1 Introduction

In this information era, how to identify a person's identity has become a key social problem. Biometrics provides a solution to this problem. The surface of hand contains many biometric traits that could be used in personal authentication systems. Some of them, such as the inner knuckle print (IKP), are relatively new and deserved intensively studying [1].

In Li et al. [2], developed an algorithm based on a  $48 \times 48$  region extracted from the finger image, which contains the line pattern of the inner middle knuckle. Ribaric et al. extracted eigenfinger features for personal identification from the

---

M. Liu · J. Yan (✉)  
College of Electronic and Information Engineering,  
Hebei University, Baoding 071002, Hebei, China  
e-mail: ch\_ao\_jun@163.com

M. Liu  
e-mail: liuming@hbu.edu.cn

whole finger image [3]. In Luo et al. [4], proposed to detect line features of the IKP by Radon transform and singular value decomposition.

In this paper, the noise and the minute lines are removed from the image based on the guided image filtering firstly. Then the thick lines are extracted from the image with a derivative edge detector. The binary line images can be matched with cross-correlation, and high verification accuracy can be obtained.

The rest of the paper is as follows: Sect. 53.2 introduces the ROI extraction method; Sect. 53.3 presents line features extracted by guided image filtering; Sect. 53.4 discusses the images registration based on cross-correlation method; Sect. 53.5 shows the experimental results and Sect. 53.6 concludes the paper.

## 53.2 Region of Interest Extraction

Images of our experiment were collected by ourselves. Because of the collected images were three fingers (middle finger, ring finger and little finger) together. We must segment the single finger and then extract the region of interest (ROI) accurately in these fingers. The ROI extraction algorithm consists of the following steps:

Step 1: Transform the original images to gray level (see Fig. 53.1a).

Step 2: Extract horizontal lines with Gabor filter and binarize the resulted image (see Fig. 53.1b). The horizontal projection histogram has three peaks (see Fig. 53.1c), corresponding to the vertical coordinates of the boundaries.

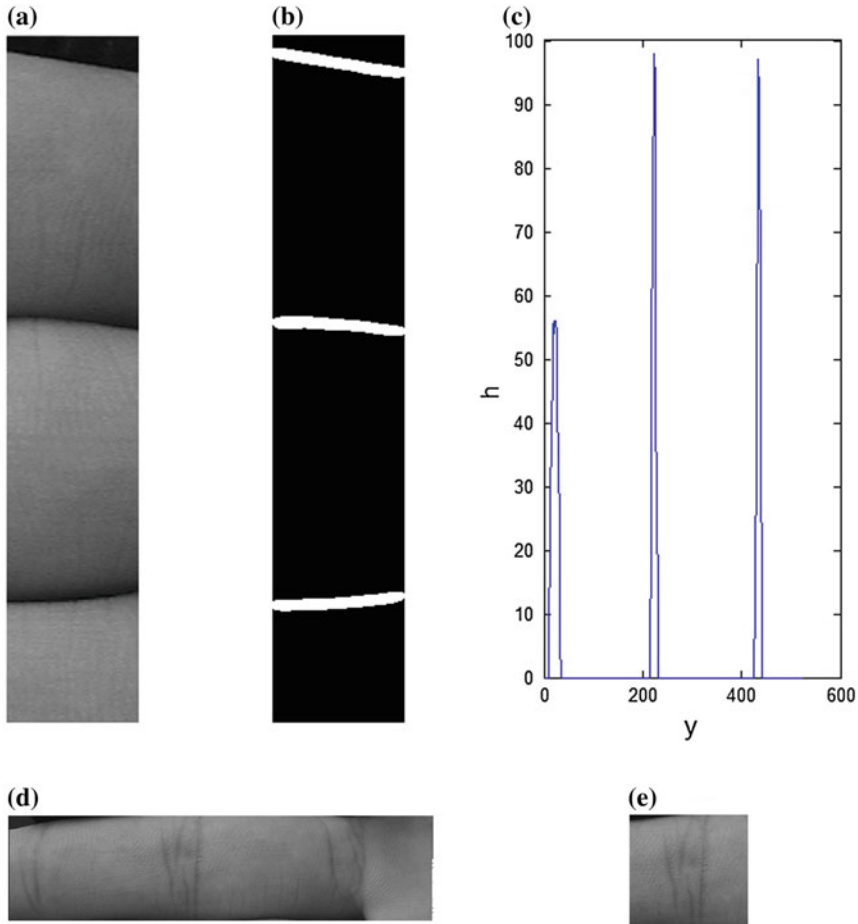
Step 3: Segment the middle finger region from the original image based on the three peaks (see Fig. 53.1d).

Step 4: Crop the ROI from the middle finger region (see Fig. 53.1e).

After the above process, ROI images are normalized to get sample images which have unified mean and variance.

## 53.3 Line Features Extracted Based on Guided Image Filtering

In this section, we will first introduce the guided image filtering method, and then discuss the derivative line detection method. The guided filter has good edge-preserving properties like the popular bilateral filter [5]. In addition, the derivative line detection method can obtain thick lines and remove noise. The combination of them can extract the line features commendably.



**Fig. 53.1** Region of interest extraction. **a** Portion of the original image. **b** Edge detection result. **c** Horizontal projection histogram. **d** Segment the middle finger region. **e** The ROI

### 53.3.1 Guided Image Filtering

In order to remove noise, we can adopt the image filtering. Guided image filtering can well remove the small noise and keep the edge features simultaneously. It includes three elements: a guidance image  $I$ , a filtering input image  $p$ , and an output image  $q$ . Among them,  $I$  and  $p$  are given beforehand, and they can be identical. Image  $q$  is considered as a linear transform of  $I$  in a window centered at the pixel  $j$ :

$$q_i = a_j I_i + b_j, \quad \forall i \in w_j \tag{53.1}$$

where  $(a_j, b_j)$  are constant in  $w_j$ . In order to get  $(a_j, b_j)$ , the following linear ridge regression model [6] is minimized in the window  $w_j$ :

$$E(a_j, b_j) = \sum_{i \in w_j} \left( (a_j I_i + b_j - p_i)^2 + \varepsilon a_j^2 \right) \tag{53.2}$$

Here  $\varepsilon$  is a parameter in order to prevent  $a_j$  excessive. The solution of Eq. (53.2) is given by:

$$a_j = \frac{\frac{1}{|w_j|} \sum_{i \in w_j} I_i p_i - \mu_j \bar{p}_j}{\sigma_j^2 + \varepsilon} \tag{53.3}$$

$$b_j = \bar{p}_j - a_j u_j \tag{53.4}$$

where  $u_j$  and  $\sigma_j^2$  are the mean and variance of  $I$  in  $w_j$ .  $\bar{p}_j = \frac{1}{|w_j|} \sum_{i \in w_j} p_i$  is the mean of  $p$  in  $w_j$  [7].

If a pixel  $i$  is existed in all the overlapping windows, the value of  $q_i$  is not identical when it is computed in different windows. So Eq. (53.1) can be rewrote by:

$$q_i = \bar{a}_i I_i + \bar{b}_i \tag{53.5}$$

where  $\bar{a}_i = \frac{1}{|w_j|} \sum_{j|i \in w_j} a_j$ ,  $\bar{b}_i = \frac{1}{|w_j|} \sum_{j|i \in w_j} b_j$  [7]. Equations (53.3)–(53.5) are the definition of guided filter.

There are two schemes to apply guided filter in our experiment. One is self-guided filter, that is to say  $I$  and  $P$  are identical. The other is mutual guided filter, that is to say  $I$  and  $P$  are different. Both of them can keep edge features and remove noise. But the self-guided filter may keep the noise by mistake. So we adopt mutual guided filter. The process is designed as follows:

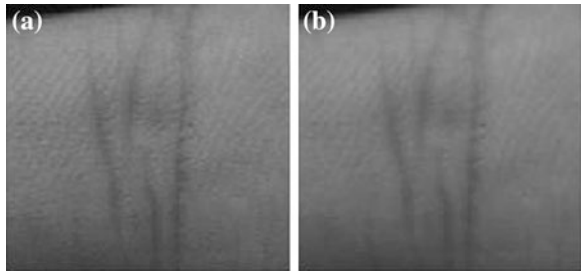
Step 1: Register two ROIs, one of which is used as input image and the other is guidance image.

Step 2: Choose a suitable window.

Step 3: Exchange the position of  $I$  and  $p$ , repeating the above experiment.

Figure 53.2 shows the effect of guided images filtering.

**Fig. 53.2** The effect of guided image filtering: **a** original image, **b** image based on guided filtering



### 53.3.2 Line Detection

After guided image filtering, we detected the IKP lines with the derivative line detection method [8]. Let  $I$  denote a discrete gray level image. Using a  $l \times m$  mean filter to smooth  $I$ , we get

$$I_1 = I * h \quad (53.6)$$

For each pixel  $(i, j)$  in  $I$ , we calculate the first order derivatives at pixels  $(i - r, j)$  and  $(i + r, j)$  along the x-axis by the following equations:

$$\left(\frac{\partial I_1}{\partial x}\right)(i - r, j) = I_1(i, j) - I_1(i - r, j) \quad (53.7)$$

$$\left(\frac{\partial I_1}{\partial x}\right)(i + r, j) = I_1(i + r, j) - I_1(i, j) \quad (53.8)$$

Then the second order derivative at pixel  $(i, j)$  can be computed by:

$$\left(\frac{\partial^2 I_1}{\partial x^2}\right)(i, j) = \left(\frac{\partial I_1}{\partial x}\right)(i + r, j) - \left(\frac{\partial I_1}{\partial x}\right)(i - r, j) \quad (53.9)$$

The horizontal lines can be obtained by looking for the zero-crossing points of the first order derivative and their strengths are the values of the corresponding second order derivative, i.e.

$$L_1(x, y) = \begin{cases} \left(\frac{\partial^2 I_1}{\partial x^2}\right)(i, j), & \text{if } I_2(i, j) \times I_3(i, j) \leq 0 \\ 0, & \text{else} \end{cases} \quad (53.10)$$

Finally, we threshold image  $L_1$  and get a binary line image  $L$ , i.e.

$$L(x, y) = \begin{cases} 1, & \text{if } L_1(x, y) > D \\ 0, & \text{else} \end{cases} \quad (53.11)$$

Figure 53.3 presents binary line image obtained by this method.

**Fig. 53.3** Binary line image obtained by derivative method



### 53.4 Registration Based-on Cross-Correlation

In our system, we adopt a cross-correlation method. The cross-correlation operator can estimate the translation displacement between the binary images accurately [9]. The main idea is: there are two images, one is matched image and the other is reference image which can be regarded as templates traverse on the matched images. Then calculate the cross-correlation value between them on all positions. The maximum of cross-correlation value is selected as the optimal position.

Let  $T(x, y)$  denote a  $M_1 \times N_1$  binary IKP line image registered in the database, and  $L(x, y)$  be a  $M_2 \times N_2$  binary IKP line image which is to be verified. Let  $M = \max\{M_1, M_2\}$ ,  $N = \max\{N_1, N_2\}$ . Then  $T(x, y)$  and  $L(x, y)$  can be enlarged to  $M \times N$  by zero-padding. The cross-correlation operation is defined by:

$$C_{T,L}(u, v) = \sum_{x=1}^M \sum_{y=1}^N T(x, y)L(x - u, y - v) \quad (53.12)$$

If  $T(x, y)$  and  $L(x, y)$  differ only by translation displacement, i.e.

$$T(x, y) = L(x - u_0, y - v_0) \quad (53.13)$$

By maximizing  $C_{T,L}(u, v)$  we can obtained the displacement  $(u_0, v_0)$ . Fourier transform operation is adopted to enforce the cross-correlation value of Eq. (53.12). So the Eq. (53.12) can be expressed as follows

$$C_{T,L}(x, y) = FT^{-1}\{FT(T(x, y))FT^*(L(x, y))\} \quad (53.14)$$

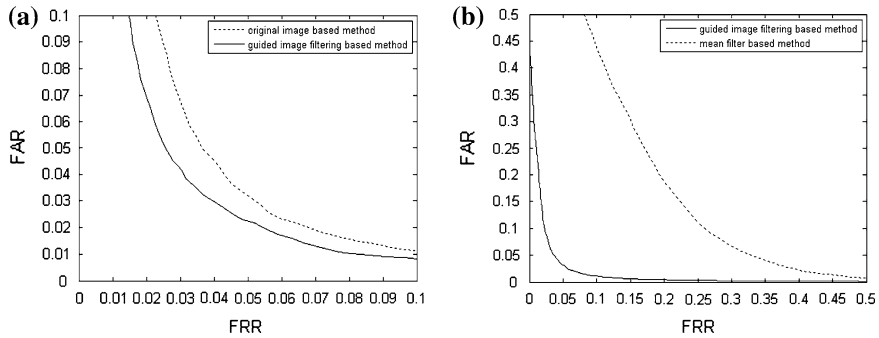
Here the FT is the Fourier transform operation and  $FT^{-1}$  is the Fourier inverse transform operation, the  $*$  is the complex conjugate operator. The position corresponding to the maximum value of  $C_{T,L}(x, y)$  shows the translation parameters registering images.

### 53.5 Experimental Results

In this paper, we use the database which were established by ourselves. There are totally 2000 images of 100 different subjects. The first 1000 samples were captured in the first library and the others in the second library.

In verification, the input IKP is known, and each IKP image is matched with the first image of all persons in the database. If two IKP images are from the same finger, we regarded it as a correct matching and called intra-class. Otherwise, inter-class means inaccurate matching. We select an image from one library, there are totally 19800 intra-class comparisons and 110000 inter-class comparisons.

The proposed method was compared with the original image based method and the mean filter based method. Figure 53.4 depicts the ROC curves obtained with the middle IKPs. The corresponding EER values are listed in Table 53.1. We can see that the proposed method outperforms all of the two conventional methods.



**Fig. 53.4** Comparison of different methods. **a** Comparison of original image based method with guided image filtering based. **b** Comparison of mean filter based method with guided image filtering based method

**Table 53.1** EER for different methods

Verification schemes	Original image based method	Mean filter based method	Guided filter based method
EER(%)	5.99	4.38	3.20

## 53.6 Conclusion

In the paper, guided image filtering is used to remove the noise from ROI, and then extract line features based on a derivative edge detector. The binary line images are matched with a cross-correlation-based image matching method. This new method can not only remove noise but also increase the accuracy in line matching. The experimental results show that the new method is suitable for personal verification.

**Acknowledgments** This work is supported by the National Natural Science Foundation of China (No. 60903089, No. 60773062, No. 61100143, No. 60801053), Scientific Research Plan Projects of Hebei Educational Bureau (No. 2008312), Beijing Natural Science Foundation (No. 4082025), Science and Technology Support Program of Hebei Province (No. 12210137).

## References

1. Jain AK, Duin RPW, Mao J (2000) Statistical pattern recognition: a review. *IEEE Trans Pattern Anal Mach Intell* 22(1):4–37
2. Li Q, Qiu ZD, Sun DM et al (2004) Personal identification using knuckle print. *Advances in Biometric Person Authentication*. In: *Proceedings of SINOBIO METRICS'04*, pp 680–689
3. Ribaric S, Fratric I (2005) A biometric identification system based on eigenpalm and eigenfinger features. *IEEE Trans Pattern Anal Mach Intell* 27(11):1698–1709
4. Luo RF, Lin TS, Wu T (2007) Personal recognition with finger crease pattern. *Opto-Electron Eng* 34(6):116–121

5. Tomasi C, Manduchi R (1998) Bilateral filtering for gray and color images. In: Proceedings of IEEE international computer vision (ICCV) conference
6. Draper N, Smith H (1981) Applied regression analysis, 2nd edn. Wiley, New York
7. He K, Sun J, Tang X (2010) Guided image filtering. *Lect Notes Comput Sci* 6331:1–14
8. Wu X, Zhang David, Wang K (2006) Palm-line extraction and matching for personal authentication. *IEEE Trans Syst Man Cybern Part A* 36(5):978–987
9. Goshtasby AA, Gage SH, Bartholic JF (1984) A two-stage cross-correlation approach to template matching. *IEEE Trans Pattern Anal Mach Intell* 6(3):374–378



# Chapter 54

## Face Recognition Using Sequence Combinatorics

Chunhui Wang, Ankang Hu and Fenglei Han

**Abstract** This paper presents a sequence similarity, called all common subsequences (ACS), for use with support vector machine (SVM) and k-nearest neighbors (kNN) to the face recognition problem. We first decompose face images as row and column sequences. Then use ACS, which compares two sequences by counting the number of occurrence of common subsequences, to measure the similarity of each pair of corresponding sequences in two images and the average of similarity of all pairs of sequences is proposed to be the similarity of two images. Experiments on four public face databases: Caltech, Jaffe, Orl and Yale databases, demonstrate that ACS can achieve higher recognition accuracy than some classic face recognition methods, e.g. 2DPCA and 2DLDA.

**Keywords** All common subsequences (ACS) · Face recognition kNN · SVM · 2DPCA · 2DLDA

### 54.1 Introduction

Face recognition is one of the hottest research topics aimed at biometric applications such as robotics, visual surveillance, human–computer interfaces etc. Comparing with other biometrics recognition, e.g. fingerprint, eye iris recognition, face recognition provides more natural means for perceptual interface without special requirements for user actions while only makes use of a wide range of inexpensive consumer cameras. However, the face recognition technic for consumer applications still remains a difficult problem. The main problem is that most of face recognition algorithms weakly perform under the some common

---

C. Wang (✉) · A. Hu · F. Han  
College of Shipbuilding Engineering, Harbin Engineering University,  
No. 145 NanTong Street, Harbin 150001, NanGang District, PRC  
e-mail: chunhui\_wang@hrbeu.edu.cn

conditions, which include the variation of facial expressions or lightening conditions, the occlusion of faces like wearing glasses or mask, the low resolution or noises of input images, and the like. The other problem is the recognition efficiency, especially when the facial database is tremendous. A good recognition algorithm should react in real time. Thereby, it is a great challenge for a face recognition algorithm to achieve high robustness and computational efficiency. Currently, main methods for face recognition include Linear Subspace Method (e.g. PCA, LDA etc.), Nonlinear Subspace Method, i.e. Kernel Method (e.g. Kernel PCA etc.), Elastic Graph Matching (e.g. DLA etc.), Neural Network-Based Method (e.g. CNN etc.), Hidden Markov Model and so on. In this work we firstly present a novel kernel method—all common subsequences (ACS)—for face recognition. ACS [1] was designed for solving the time series problem. The main concept of ACS is to measure the similarity of two sequences by counting the number of all common subsequences of these two sequences. For the use of ACS for images, we need to translate images to sets of sequences, and then individually use ACS for each pairs of sequences to compute the similarity, in the end average the results as the similarity between images.

The rest of the paper is organized as follows. In Sect. 54.2 propose the methodology of measuring similarity of images using all common subsequences. The experimental results of face recognition are presented in Sect. 54.3. This paper is concluded in Sect. 54.4.

## 54.2 All Common Subsequences

*All common subsequences* was firstly proposed in [1] for the purpose of measuring the similarity of time series. Then it was extensively studied in [2] as a problem of computer science. Later ACS was proved to be a valid kernel [3], thereby ACS could also be studied in kernel machine.

Let  $\Sigma$  be a finite alphabet. An  $n$ -long sequence  $t$  is an ordered set  $\{t_1, \dots, t_n\}$ , where  $t_i \in \Sigma$  is the  $i$ th element in sequence  $t$ ,  $1 \leq i \leq n$ . An empty sequence is denoted by  $\varepsilon$ , whose length is 0. Let  $u$  be a sequence. If there exist indices  $i = (i_1, \dots, i_{|u|})$ , with  $1 \leq i_1 < \dots < i_{|u|} \leq |t|$ , such that  $u_j = t_{i_j}$ , for  $j = 1, \dots, |u|$ , then we say  $u$  is a subsequence of  $t$  (denoted by  $u \leq t$  or  $u = t(i)$ ). We denote by  $\Sigma^n$  the set of all finite sequences of length  $n$ , and by  $\Sigma^*$  the set of all sequences

$$\Sigma^* = \bigcup_{n=0}^{\infty} \Sigma^n$$

Formally, ACS could be described as

**Definition 1** (*All common subsequences*) Let  $I$  be a feature (subsequence) space of sequences set  $S$ , lets,  $t$  be sequences, let

$$\phi_u(s) = \begin{cases} 1 & |\{i : u = s(i)\}| > 0, u \in I, \\ 0 & \text{otherwise.} \end{cases}$$

Then, ACS can be defined as an inner product of vectors of  $\phi_u(s)$ :

$$acs(s, t) = \langle \phi_u(s), \phi(t) \rangle = \sum_{u \in I} \phi_u(s) \phi_u(t) \tag{54.1}$$

Dynamic approach is adopted for calculating the number of ACS. Lemma 1 implies quadratic operations.

**Lemma 1** *Let  $s$  and  $t$  be finite, nonempty sequences over  $\Sigma$  with lengths  $|s| = m$  and  $|t| = n$  respectively. For each  $p \in \Sigma$  let  $r(s, p) := \max\{i : s_i = p\}$  with  $r(s, b) := 0$  if  $p > s$ . For brevity, we set  $s^i := s(1 : i)$ ,  $r_s := r(s^{m-1}, s_m)$  and  $r_t := r(t^{n-1}, t_n)$ . Then*

$$acs(s, t) = \begin{cases} acs(s^{m-1}, t) + acs(s, t^{n-1}) - acs(s^{m-1}, t^{n-1}) & \text{if } s_m \neq t_n, \\ acs(s^{m-1}, t^{n-1}) \times 2 - acs(s^{r_s-1}, t^{r_t-1}) & \text{if } s_m = t_n, 0 < r_s < m \text{ and } 0 < r_t < n, \\ acs(s^{m-1}, t^{n-1}) \times 2 & \text{otherwise.} \end{cases} \tag{54.2}$$

### 54.2.1 All Common Subsequences for Images

The visual appearance of objects in computing science is represented by digital images, which have a finite set of digital values, called pixels. Let  $I$  be an image with  $m \times n$  pixels. Since each pixel of the image carries a single integer value (0 1 for binary and 0 255 for grayscale image), the image could be described by an matrix with integer value entries, denoted also by  $I$ .

Let  $R^i, C^j$  be the  $i$ th row,  $j$ th column of the image  $I$ , respectively. Then  $I$  could be deemed as the orderly combination of its rows  $I = [R^1; R^2; \dots; R^m]$  or the orderly combination of its columns  $I = [C^1; C^2; \dots; C^n]$ , where each row is a sequence with length  $|R^i| = n$  and each column with length  $|C^j| = m$ . Thereby we transformed an image to a set of sequences:  $I = \{R^1, R^2, \dots, R^m, C^1, C^2, \dots, C^n\}$ . For convenience, we denote  $R^i$  by  $I^i, i = 1 \dots m$ , and denote  $C^j$  by  $I^{m+j}, j = 1 \dots n$ . Then the sequences set of image  $I$  is

$$I = \bigcup_{i=1}^{m+n} I^i.$$

**Definition 2** The similarity of two images  $I, J$  with the same size  $m \times n$  is the sum of number of all common subsequences of corresponding rows and columns of images  $I$  and  $J$ :

$$acsi(I, J) = \sum_{i=1}^{m+n} acs(I^i, J^i) \quad (54.3)$$

Formally, ACS should be normalized:

$$\widehat{acsi}(I, J) = \frac{1}{m+n} \sum_{i=1}^{m+n} \frac{acs(I^i, J^i)}{\sqrt{acs(I^i, I^i)acs(J^i, J^i)}} \quad (54.4)$$

The computational efficiency of ACS of two sequences with length  $m$  and  $n$  is  $O(mn)$ . So the computational efficiency of ACS of two images with size  $m \times n$  is  $O(m^2n + mn^2) = O(\max(m^2n, mn^2))$ .

### 54.3 Face Recognition Experiments

Before calculating the value of ACS of images, the original images should be preprocessed. Generally, image preprocessing consists of the following procedures:

- *Cropping*—Some images may contain wide background, which should be cropped off.
- *Converting*—Original images sometimes are color, while ACS only concern the luminance of images. So we need to convert color images to grayscale by eliminating the hue and saturation information while retaining the luminance.
- *Compensation and Equalization*—Face under different lightening conditions shows appearances with different or unbalanced luminance values, while ACS directly compares the gray value of images. So we should make sure that the gray value of the same face under various lightening conditions is invariant. In this work we compensate the face illumination by the technic based on wavelet transform [4], and equalize the pixels of each gray value of images using histogram equalization.
- *Resizing*—Images from datasets have high (spatial) resolution (e.g.  $256 \times 256$  or higher). But in experiments, the time for computing ACS of images is proportionate to the cubic of image size. So we should make the resolution of images as low as possible.
- *Rescaling*—Grayscale images often possess high gray-level resolution (e.g. 8-bit), which may cause ACS not robust, i.e. make ACS sensitive to the noise or slight changes of luminance. Hence it is required to adjust high gray-level resolution to lower. Experimental results suggest that under 1-bit gray-level the

recognition performs best. For pixel with gray value  $p$ , we can decrease the gray-level resolution in this way:  $\tilde{p} = \left\lfloor \frac{p}{\mu} \right\rfloor$ , where  $\tilde{p}$  is a new gray value and  $\mu \in [100, 140]$  is a rescaling factor.

Rescaling factor  $\mu$  is a key parameter of ACS for face recognition, for the value of significantly affects the recognition accuracy. As for 2DPCA and 2DLDA, the rescaling step is not demanded.

### 54.3.1 Face Recognition on Public Databases

We use ACS with SVM [5] and  $k$ NN for face recognition on four public databases: Caltech [6], Jaffe [7], Orl [8] and Yale [9] face databases. In order to make the experiments comparable, two classic face recognition methods are adopted in experiments: 2DPCA [10] and 2DLDA [11].

The penalty parameters of SVM is set to  $C = 10$ . The parameter  $k$  of  $k$ NN is tuned to the value which makes the recognition get the highest accuracy. The preprocessing steps, including *cropping*, *converting*, *compensation*, *equalization*, *resizing*, *rescaling*, are orderly carried out if necessary. The images are resized to  $32 \times 32$ . The rescaling factor  $\mu$  is tuned carefully to achieve the best recognition results. All experiments, if not specialized, are performed with leave-one-out strategy.

From Caltech database we choose 395 images for face recognition experiment, which consist of 19 subjects (each one has 18 25 images). All images have various backgrounds. So image preprocessing steps include *cropping*, *converting*, *equalization*, *resizing*, *rescaling*.

Jaffe database contains 213 images of 7 facial expressions (anger, disgust, fear, happiness, neutral, sadness, surprise) posed by 10 Japanese female expressers. Image preprocessing steps include *cropping*, *converting*, *equalization*, *resizing*, *rescaling*.

Orl database consists of 40 distinct subjects, each of which has ten different images. All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for slight side movement). The image preprocessing steps only include *converting*, *resizing*, *rescaling*.

Yale face database includes 15 subjects with 11 images per subject: center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised, and winking. So the unbalanced light exists in this database. Hence the image preprocessing steps include *cropping*, *converting*, *compensation*, *equalization*, *resizing*, *rescaling*.

Figures 54.1, 54.2, 54.3, and 54.4 depict some faces from Caltech, Jaffe, Orl, Yale databases (the upper photos) and the preprocessed ones (the nether photos), respectively.



Fig. 54.1 Some faces from Caltech database. *Upper* raw faces; *nether* preprocessed faces

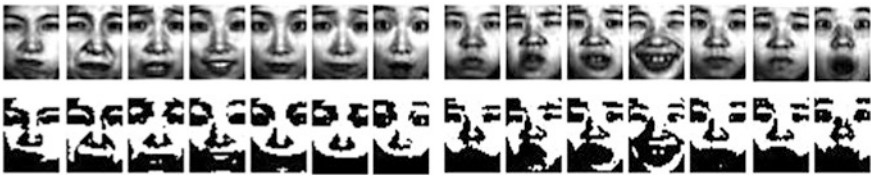


Fig. 54.2 Some faces from Jaffe database. *Upper* raw faces; *nether* preprocessed faces

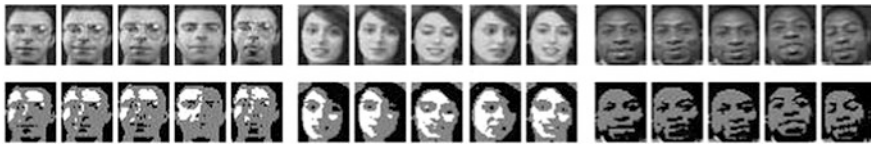


Fig. 54.3 Some faces from Orl database. *Upper* raw faces; *nether* preprocessed faces

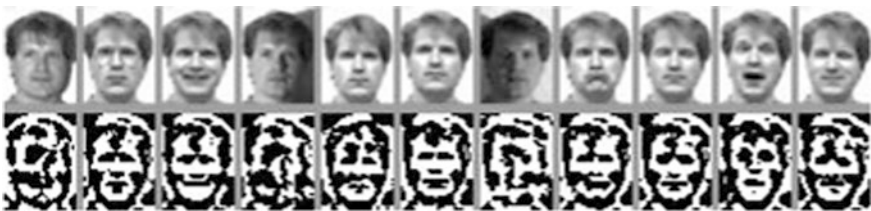


Fig. 54.4 Some faces from Yale database. *Upper* raw faces; *nether* preprocessed faces

The experimental results of face recognition are shown in Table 54.1. From the table we find that

- ACS with  $k$ NN outperforms other methods in most cases;
- Algorithms with  $k$ NN in some cases perform better than that with SVM.

**Table 54.1** Face recognition accuracy (%) on four databases

Method	Caltech	Jaffe	Orl	Yale
acs_svm	99.24	100.0	100.0	96.97
2dpca_svm	99.75	100.0	98.25	97.58
2llda_svm	99.49	100.0	98.5	98.18
acs_knn	98.99	100.0	100.0	99.39
2dpca_knn	97.97	99.06	98.25	98.18
2llda_knn	98.99	100.0	97.5	96.97

## 54.4 Conclusion

In this work we present all common subsequences (ACS) as the kernel function (similarity method) to scatter images into a Euclidean space, then classify them in SVM and  $k$ NN classifiers to solve the face recognition problem. Firstly images are decomposed as row and column pixel sequences. Secondly ACS which compares two sequences by counting the number of occurrence of common subsequences is adopted to measure the similarity of each pair of corresponding sequences in two images and the average of similarity of all pairs of sequences is proposed to be the similarity of two images. Finally images are classified in  $k$ NN and SVM. We performed experiments on 4 public face databases: Caltech, Jaffe, Orl and Yale databases, and demonstrated that ACS is a good kernel function (similarity method) that outperforms 2DPCA and 2DLDA.

## References

1. Wang H (2007) All common subsequences. In: Proceedings of the 20th international joint conference on artificial intelligence (IJCAI'07), Hyderabad, India, pp 635–640
2. Elzinga C, Rahmann S, Wang H (2008) Algorithms for subsequence combinatorics. *Theoret Comput Sci* 3(409):394–404
3. Guo Z, Wang H, Lin Z, Guo X (2010) A study of all common subsequences in kernel machine. In: Proceedings of the international conference on machine learning and cybernetics (ICMLC) 2010, Qingdao, China, pp 1763–1766
4. Nie XF, Tan ZF, Guo J (2008) Face illumination compensation based on wavelet transform. *Opt Precis Eng* 16(1):150–155
5. Canu S, Grandvalet Y, Guigue V, Rakotomamonjy A (2005) SVM and kernel methods matlab toolbox. Perception Systems et Information, INSA de Rouen, Rouen, France
6. Fei-Fei L, Fergus R, Perona P (2004) Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. In: IEEE CVPR 2004, workshop on generative-model based vision
7. Lyons MJ, Akamatsu S, Kamachi M, Gyoba J (1998) Coding facial expressions with gabor wavelets. In: Proceedings of the international conference on automatic face and gesture recognition, pp 200–205
8. Samaria F, Harter A (1994) Parameterization of a stochastic model for human face identification. In: Proceedings of the IEEE workshop on applications of computer vision, pp 138–142

9. Belhumeur PN (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans PAMI* 19(7):711–720
10. Yang J, Zhang D, Frangi AF (2004) Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Trans Pattern Anal Mach Intell* 1(26):131–137
11. Li M, Yuan B (2005) 2D-LDA: a statistical linear discriminant analysis for image matrix. *Pattern Recogn Lett* 5(26):527–532



# Chapter 55

## Distinction of Breast Tissues Based on Segmented Integral Area of Frequency-Resistance Curves

Chao Wang, Yiming Wei and Ruifeng Bai

**Abstract** Breast cancer seriously endangers the health of women, which makes intra-operative assessment of cancer focus have vital significance. The information of bioelectrical impedance has unique ability to distinguish cancerous and normal tissue, and can provide basis for intra-operative assessment of cancer focus. In order to achieve accurate measurement, a measurement system is established composed of the impedance analyzer and probe with optimized electrode. Segmented integral area is regarded as characteristic parameter to reflect the over all trend. To utilize the advantages of different frequency-resistance curves, BP neural network is finally selected and good-training neural networks are integrated to make the final decision. The result indicates that the characteristic parameter selected can reflect differences of tissues and the integrated BP neural network has better performance than single neural network.

**Keywords** Breast cancer · Biological impedance measurement technology (BIMT) · Characteristic parameter · BP neural network · Frequency-resistance curves

### 55.1 Introduction

Breast cancer is the most common type of cancer among women, the incidence rate of which has an upward trend in recent years. According to the GLOBOCAN 2008 distributed by the international cancer research center of the world health organization, the incidence and mortality of breast cancer, respectively, accounted

---

C. Wang (✉) · Y. Wei · R. Bai  
Tianjin Key laboratory of process Measurement and Control,  
School of Electrical Engineering and Automation, Tianjin University,  
No. 92 Weijin Road, Tianjin 300072, China  
e-mail: wangchao@tju.edu.cn

for 10.9 and 6.1 % of the total number of death due to cancer in 2008. Predicted by the research, breast cancer may have the highest incidence rate among all cancer in five years [1].

Bioelectrical impedance measurement technology is a non-invasive detection method, which is based on the theory that the breast tumor and normal tissues have significant differences in electrical properties [2–5], and can provide powerful basis for identifying breast cancer focus. To achieve the target accurately and quickly, the probe contacting the tissue should satisfy the measuring requirement, and characteristic parameter with high partition degree is also needed. Probe with four-electrode is the main means for biological impedance measurement at low frequencies, which can reduce the effect of contact impedance. The extraction of characteristic parameter has been exploring for a long time. A lot of researchers attempt to find properties from different frequency bands. Several parameters related to electrical and impedance characteristic have been used to distinguish tissues, such as admittivity, real part and imaginary part of complex impedance [6–8]. There are also some researchers trying to deal with the characteristic parameter using mathematical method like linear algebra [9], combination of ANN, SVM and GMM [10].

In this paper, the biological impedance information is detected by measurement system and the data obtained is painted as frequency-resistance curves. Characteristic parameter which can reflect the over all trend of frequency-resistance curves has been extracted. BP neural network is built up and several networks are integrated to improve the accuracy of distinction.

## 55.2 Measurement System

In order to achieve accurate measurement at different frequencies, the impedance analyzer (Agilent 4294A) and a probe with optimized electrode are used to make up the measurement system.

The optimized arc-shaped electrode is shown in Fig. 55.1, which has one pair of exciting electrode and one pair of measuring electrode [11, 12]. The excision tissues are chosen to be the subjects in this paper to reduce the influence of human body. When measuring the subject, the electrode is placed on the surface of the tissue. The exciting electrode provides small alternating current signals, and meanwhile, the measuring electrode detects the voltage signals from the subject. The sensitive area of electrode is of 1 cm wide and 0.2 cm underneath the electrode. The purpose of impedance spectra measurement is to judge whether the focus under the sensitive area has been completely excised, if not, further excision is needed.

The measurement system is shown in Fig. 55.2. The frequency range of the excitation source is 40–1.5 MHz. The impedance of tissues and information related to the characteristic parameter are calculated by the host computer after receiving the data from the impedance analyzer.

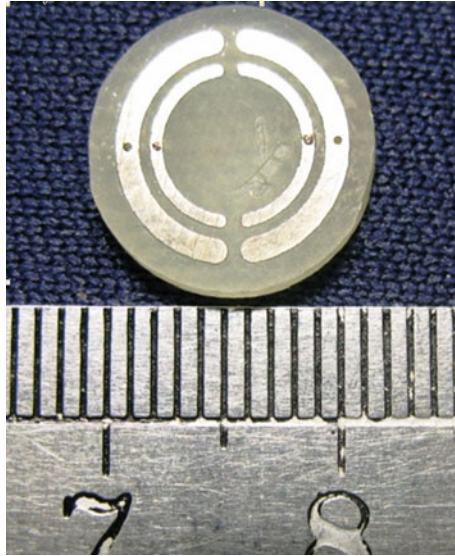


Fig. 55.1 Arc-shaped electrode

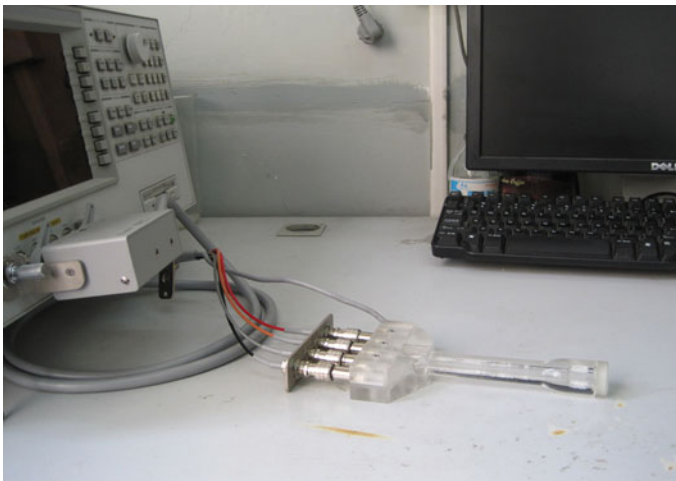
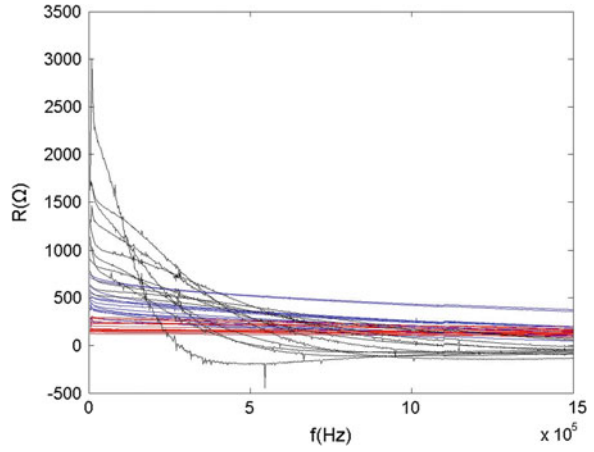


Fig. 55.2 Measurement system

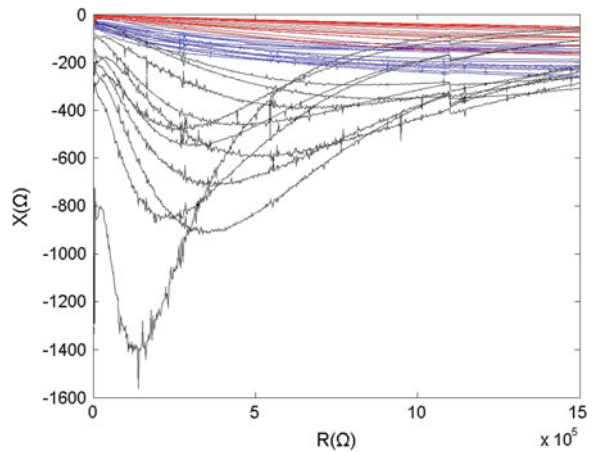
### 55.3 Frequency-Resistance Curves Analysis and Characteristic Parameter Extraction

Every data point is recorded every 3 kHz, and there are 800 points for each sample. The frequency-real part (f-R), frequency-imaginary part (f-X) and real-imaginary (R-X) part curves are shown in Figs. 55.3, 55.4 and 55.5.

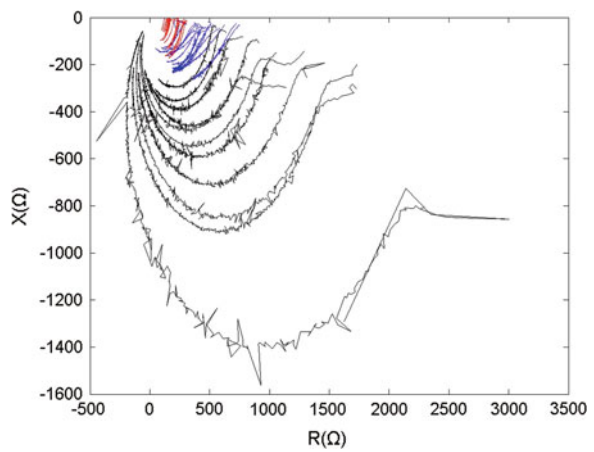
**Fig. 55.3** Frequency-real part curves of three tissues. *Blue* curves stand for cancer, *red* stand for gland and *black* stand for adipose



**Fig. 55.4** Frequency-imaginary part curves of three tissues



**Fig. 55.5** Real-imaginary part curves of three tissues



The curve trend of each tissue differs from one another, but it still follows certain rules. Curves of gland are the gentlest, while adipose changes greatly, and cancer varies between them; the curves of adipose have drastic changes, while other two change smoothly but still have differences. It can be seen that, for the same tissue, with the change of horizontal axis, the value of longitudinal axis varies greatly, but the curve trend remains constant on the whole. The curves of different tissues show different regular patterns, while curves of same tissue have good reproducibility. So it's possible to distinguish tissues using the trend of frequency-resistance curves.

The integral area of the curve changes with the trend, so the integral area of different tissues should be different from each other owing to the trend. Since the integral area can reflect the change of curve trend, it can be chosen as the characteristic parameter.

### 55.3.1 Preliminary Study of Distinction Based on Segmented Integral Area

To reflect the regular pattern of the curve change using the integral area, every 20 data points on the curve become one group according to the change of horizontal axis, so that all of the points of each curve are divided into 40 cantlets. Integrating the data points using the trapezoidal integration method, 40 segmented integral areas can be got for each sample. To analyze the distinction and concentration degree of segmented integral area, the average value and standard deviation of segmented integral areas extracted from f-R, f-X and R-X curves are calculated and shown in Tables 55.1, 55.2 and 55.3. Since the numbers of f-R and f-X is too large to be seen clearly, they have been divided by 10<sup>6</sup> first.

The average value and standard deviation of gland is the smallest among the three tissues of the three curves, while adipose is the largest and cancer in the middle. This phenomenon is the same with the curve trend analyzed above, and it also proves that segmented integral areas can reflect the curve trend.

The change intervals of cancer and adipose of f-R curve are almost coincident, but gland is prominent. For the average value of f-X curve, the change interval of gland has no coincidence with others', but cancer and adipose have part of the same interval. The standard deviation of gland is the smallest, which illustrates the values of gland are the most concentrated. As for the R-X curve, the change interval of adipose is obviously different from other two, but the cancer and gland have the same part.

**Table 55.1** Average value and standard deviation of f-R

Range of variation	Cancer	Gland	Adipose
Average value	8.13–20.2	4.14–8.63	8.42–21.9
Standard deviation	1.223–2.912	0.167–0.822	3.123–26.508

**Table 55.2** Average value and standard deviation of f-X

Range of variation	Cancer	Gland	Adipose
Average value	-9.09 to -5.70	-3.97 to -1.04	-19.11 to -8.57
Standard deviation	0.712-2.063	0.496-1.729	2.606-15.188

**Table 55.3** Average value and standard deviation of R-X

Range of variation	Cancer	Gland	Adipose
Average value	-1580.68 to -423.98	-432.806 to -14.338	-84432.2 to -2675.96
Standard deviation	110.418-776.742	8.415-214.138	1281.932-1.918*10 <sup>5</sup>

It can be concluded that segmented integral area of f-X curve has a good ability in distinguishing gland, but can not recognize cancer and adipose. Adipose can be distinguished using R-X curve in numerical, even though the curves are scattered, but cancer and gland can not.

The frequency-resistance curves have particular ability in distinguishing certain tissue, but they can not distinguish three tissues separately. Segmented integral areas can not be used as characteristic parameter directly. So there is a strong need for a new intelligent method which can combine the advantages of three curves and be better suited to identify tissues.

### 55.3.2 BP Neural Network Integration in Distinction

To combine the advantages of frequency-resistance curves, the integrated BP neural network consists of three parts, f-R, f-X, R-X networks. The identification process is shown in Fig. 55.6. Characteristic parameters extracted from three curves are regarded as inputs of the three parts which trained separately. Each BP neural network of these parts is designed as three-layer, which contains 40 input-layer nodes, 8 hidden-layer nodes and only one output-layer node. The output values, 0, 1 and -1, represent gland, breast cancer and adipose respectively. To avoid the situation of output neuron saturation caused by large input, all of the input data are normalized by the function and vary within the range from -1 to 1. The final output is decided after integrating and judging the results from the three parts. The judgment standard can be summarized as: if two or more results are the same, then the final output is the same result; otherwise if all of the results are different, it outputs "Can not identify".

The results are integrated and judged according to certain rules for the final output. The training set contains 80 samples, including cancer 30, gland 30 and adipose 20. 60 samples are tested, including cancer 20, gland 22 and adipose 18. Part of the test results of f-R, f-X, R-X curves based on segmented integral areas are shown in Tables 55.4, 55.5, 55.6.

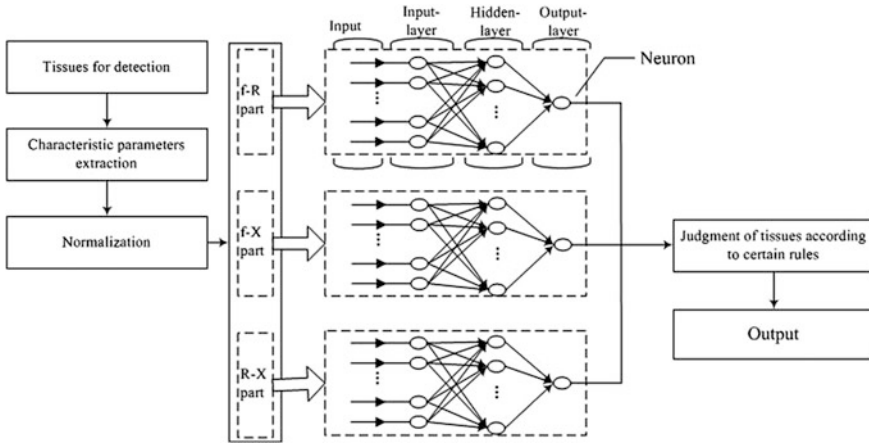


Fig. 55.6 Identification process of the integrated BP neural network

Table 55.4 Distinction result of f-R curve based on segmented integral area

Accuracy	Net1 (%)	Net2 (%)	Net3 (%)	Net4 (%)	Net5 (%)	Net6 (%)	Net7 (%)	Net8 (%)
Cancer	56.7	63.3	53.3	60	63.3	66.7	76.7	56.7
Gland	73.3	66.7	76.7	83.3	63.3	73.3	70	66.7
Adipose	61.1	55.6	66.7	72.2	66.7	77.8	72.2	66.7

Table 55.5 Distinction result of f-X curve based on segmented integral area

Accuracy	Net1 (%)	Net2 (%)	Net3 (%)	Net4 (%)	Net5 (%)	Net6 (%)	Net7 (%)	Net8 (%)
Cancer	55	85	75	60	90	90	80	80
Gland	90.9	95.5	77.3	81.8	86.4	72.7	77.3	90.9
Adipose	77.8	83.3	72.2	77.8	83.3	55.6	88.9	88.9

Table 55.6 Distinction result of R-X curve based on segmented integral area

Accuracy	Net1 (%)	Net2 (%)	Net3 (%)	Net4 (%)	Net5 (%)	Net6 (%)	Net7 (%)	Net8 (%)
Cancer	85	75	65	85	70	55	80	65
Gland	90.9	86.4	81.8	90.9	63.4	68.2	86.4	63.4
Adipose	94.4	83.3	77.8	88.9	77.8	88.9	94.4	83.3

As can be seen, single network of f-R and f-X curve can identify gland effectively in most of the time, while that of R-X curve has a good performance in identifying adipose, but the result is not good enough confronting three tissues.

**Table 55.7** Distinction result of integrated BP neural network

	Cancer (%)	Gland (%)	Adipose (%)
Accuracy	86.7	93.3	90

It can be seen from the results shown in Table 55.7 that integrated BP neural network is good at distinguishing three tissues. Using the complementary relationship between these three kinds of curves, integrated BP neural network can combine their advantages, and has better performance than single networks when dealing with three tissues.

## 55.4 Conclusion

Through detecting the impedance spectrum, f-R, f-X and R-X curves are painted. After analyzing the curve changing rule, segmented integral area is selected to show the over all trend. It's found in the preliminary study of distinction that these three curves are capable of distinguishing certain tissue, but not all tissues. Complementing the advantages of three curves, BP neural network integrates good-training networks and has better performance in distinction. With constraints of samples, more work is needed in the future.

**Acknowledgments** This work is supported by National Natural Science Foundation of China under Grant 50937005.

## References

1. Ferlay J, Shin HR, Bray F et al (2010) GLOBOCAN 2008 v1.2, cancer incidence and mortality worldwide: IARC CancerBase No. 10. Lyon: international agency for research on cancer. <http://globocan.iarc.fr>
2. Jossinet J (1996) Variability of impedivity in normal and pathological breast tissue. *Med Biol Eng Comput* 34(5):346–350
3. Jossinet J (1998) The impedivity of freshly excised human breast tissue. *Physiol Meas* 19(1):61–75
4. Surowiec AJ, Barr JB, Stuchly SS et al (1988) Dielectric properties of breast carcinoma and the surrounding tissues. *IEEE Trans Biomed Eng* 35(4):257–263
5. Chauveau N, Hamzaoui L, Rochaix P et al (1999) Exvivo discrimination between normal and pathological tissues in human breast surgical biopsies using bio-impedance spectroscopy. *Ann NY Acad Sci* 873:42–52
6. Jossinet J, Lavandier B (1998) The discrimination of excised cancerous breast tissue samples using impedance spectroscopy. *Bioelectrochem Bioenerg* 45:161–167
7. Kim BS, Isaacson D, Xia H (2007) A method for analyzing electrical impedance spectroscopy data from breast cancer patients. *Physiol Meas* 28:237–246
8. Liao Q, Xu Y, Dong X (2008) Study on impedance spectroscopic and pathological properties of breast carcinoma and surrounding tissues. *Biomed Eng Clin Med* 12(4):263–266 (in Chinese)



9. Lederman D, Zheng B, Wang X (2011) Improving breast cancer risk stratification using resonance-frequency electrical impedance spectroscopy through fusion of multiple classifiers. *Ann Biomed Eng* 39(3):931–945
10. Laufer S, Solomon SB, Rubinsky B (2012) Tissue characterization using electrical impedance spectroscopy data: a linear algebra approach. *Physiol Meas* 33:997–1013
11. Wang C, Xiao Y, Du D (2008) Electrode and measuring method for intraoperative margin. *J Tianjin Univ* 41(3):321–325 (in Chinese)
12. Wang C, Du D, Xiao Y (2008) Simulation method of electrode structure design for intraoperative margin assessment of breast cancer focus. In: IEEE international conference on machine learning and cybernetics (ICMLC2008), 2008, July, pp 2369–2372

# Chapter 56

## Vehicle Discrimination Using a Combined Multiple Features Based on Vehicle Face

Yingnan Wang, Hong Li, Clement Kipkorir Kirui and Wei Zhang

**Abstract** In this paper, a new method for vehicle discrimination and recognition on the basis of combination of multiple features based vehicle face is proposed. The color difference, vehicle face pattern difference and logo matching degree are getting together to improve the performance of vehicle discrimination. This method is assessed on a set of 200 images that belong to five distinctive vehicle manufacturers. A series of experiments are conducted, splitting the 300 pairs of images to a training set and a test set, respectively. It is shown that the enhanced feature combination approach (CMN) proposed in this paper boosts the recognition accuracy compared with the CM and CN method. The reported results indicate a high classification rate in similar or different vehicles and a fast processing time, making it suitable for real-time applications.

**Keywords** Vehicle recognition · Vehicle classification · Image matching · ITS

### 56.1 Introduction

The vehicle identification technology includes many aspects such as the license plate recognition, vehicle-logo recognition, vehicle-type classification, etc [1, 2]. However, by only depending on the traditional vehicle-type classification technology

---

Y. Wang (✉)

Central South University, Room 314, Comprehensive Experiment Building, Railway Campus, Changsha 410012, China  
e-mail: wyntabby@126.com

Y. Wang · H. Li · C. K. Kirui

School of Information Science and Engineering, Central South University, Changsha 410012, China

W. Zhang

School of Architecture, Hunan University, Changsha 410082, China

[3–5], it cannot identify other important vehicle characteristics or features. Therefore, in addition to the recognition technology which is already in use, other characteristics of vehicles such as vehicles face, logo, body color, etc., are needed to identify specific type of vehicle. Because of the above features are mainly included in vehicle face, the method of vehicle discrimination based on vehicle face studied in this paper has broad application prospects and urgent needs. It could be widely used in vehicle parking management, image matching, image retrieval, video retrieval, etc.

There are a few solutions for vehicle recognition on the basis of features in vehicle face proposed in the literature. Wang et al. [6] proposed a real-time vehicle classification scheme based on eigenface. Zhang and Zhao [7] compared random subspace ensemble of neural network classifiers with several different machine learning algorithms on the basis of the features in vehicle face extracted from Pyramid Histogram of Oriented Gradients and Curvelet Transform. What is more, there are other features are extracted, such as vehicle color and shapeme histogram projection is proposed in [8, 9].

Nevertheless, the existing methods are using the vehicle face as a whole, without considering the local feature, may lead to biggish error or classifying different brands to one class just because of the similar vehicle face pattern. Therefore, a few of recent works attempted to achieve vehicle manufacturer recognition on the basis of vehicle logo. Petrovic and Cootes [10] presented an interesting approach for vehicle manufacturer recognition from frontal-view vehicle images that displayed a 93 % recognition accuracy. Apostolos et al. [11] deals with a vehicle logo recognition problem of frontal-view images, proposing a new algorithm on the basis of an enhanced scale-invariant feature transform (SIFT) based feature-matching scheme with almost 90 % recognition rate.

For distinguish the specific vehicle model, this paper proposed a new method on the basis of combining several different features extracted from vehicle face. As a result, a composite feature description from vehicle color, vehicle face pattern and logo matching (CMN[c, m, n]) can further increase the accuracy of recognition by using their difference to compose a three-dimension feature vector. In this way, not only the vehicle manufacturer could be identified, but also the specific style of vehicle could be distinguished.

## 56.2 Vehicle Face Detection

Before feature extraction of the vehicle face, it is need to detect and segment the key feature area. The result of vehicle face detection and segmentation is a rectangular area that includes vehicle logo and radiator grill, which has abundant texture information that can be used as the important basis for vehicle identification (see Fig. 56.1).

According to a priori knowledge, such a vehicle face is located just above the license plate. So a rough location of the vehicle face could be achieved depending

Fig. 56.1 Vehicle face



on the position of the license plate and a priori knowledge. It is realizable to precisely locate the vehicle face depending on the bump location which obtained through the horizontal gradient projection. Based on the morphology and edge detection algorithm, this paper locates a license plate area, and then uses prior knowledge and horizontal gradient projection to achieve the vehicle face location.

### 56.3 Feature Descriptions

#### 56.3.1 Vehicle Color

The vehicle color as an important auxiliary feature of the vehicle identification process, describes the image regional features from the overall situation angle. It is invariant to rotation, translation, and scale variation between images. Based on the vehicle face detection, this paper extracts a rectangular area above vehicle face as the color area for the following operation (see Fig. 56.2).

In order to compare the level of similarity between a pair of vehicle images, basing on the pretreatment of color area extraction, the cumulative histogram of color is respectively calculated in RGB color space to describe the color distribution characteristics. This paper only takes the peak value of each color histogram named  $c_1$  and  $c_2$  as color features. The cumulative color histogram is calculated based on the pair of color areas. While taking the color difference as measure of the similarity, the smaller the color difference  $c$  ( $c = |c_1 - c_2|$ ), the more similar color of the images is.

#### 56.3.2 Vehicle Face Pattern

Through the observation of numerous vehicle images, the same brand automobile also has different forms and size of radiator grill. In order to distinguish the various patterns of vehicle face, this paper adopts a compared method to contrast the two vehicle face images.

According to the projection algorithm and Euclidean distance algorithm, the difference values of horizontal projection and vertical projection of this pair of

Fig. 56.2 Color area



vehicle face images are calculated as the vehicle face pattern features, which are particularly appropriate for images rich with edges. The difference values  $m_1$  and  $m_2$  between two vehicle faces are calculated now. The two difference values could be combined to a difference value as  $m$  by adding a weight value to each one (Formula 56.1). In this paper,  $a$  is set to 0.5 and  $b$  is set to 2.

$$m = \frac{\alpha \times m_1 + (1 - \alpha) \times m_2}{\beta}. \quad (56.1)$$

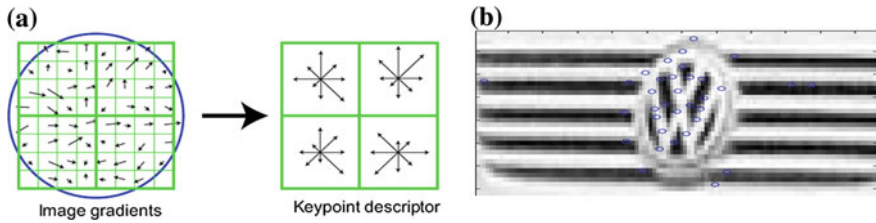
### 56.3.3 Vehicle Logo Matching

Because the above-mentioned features mainly describe image features on global, without considering the local feature, this way may lead to big error of vehicle recognition and matching. It is worth noting that vehicle logo has the most abundant texture and edge information in vehicle face. According to the similar vehicles must have the same logos, the vehicle manufacturer could be distinguished for vehicle discrimination in this paper through vehicle logo matching on the basis of SIFT method.

SIFT proposed by Lowe has a strong adaptability of feature point extraction algorithms. It exploits the idea of replacing images by a set of scale and orientation-invariant feature descriptors. To achieve the feature point extraction and matching, this paper takes the use of the scale invariance of the SIFT algorithm. In order to accomplish feature matching of two vehicle face images, the feature points  $n_1$  and  $n_2$  are detected, as well as  $n$ —the matching degree of two images.

#### 56.3.3.1 Feature Detection and Description

The invariant features are detected and extracted by exploring the scale-space structure of each image. Features are localized and filtered, keeping only those that are likely to remain stable over affine transformations, have adequate contrast, and are not along edges. The presence of keypoints not lying in the edges and having adequate contrast is ensured with the appropriate selection of two parameters  $h$  and  $\varepsilon$  in the Difference of Gaussians (DoG) function, as described analytically in [11]. Then the feature points are filtered and positioned, and its main direction is distributed, in order to get the corresponding feature descriptor (Formula 56.2). As a result, the feature descriptor is created by sampling the magnitudes and orientations of the image gradient in a patch around the detected feature, resulting in a 128-D vector (Fig. 56.3a) of direction histograms, also make these points to a certain extent with the scaling and rotation invariance. The result is presented in Fig. 56.3b.



**Fig. 56.3** Feature detection. **a** A result of the feature descriptor. **b** A result of the feature detection

$$\begin{aligned}
 m(x, y) &= \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2} \\
 \theta(x, y) &= \arctan \frac{L(x + 1, y) - L(x - 1, y)}{L(x, y + 1) - L(x, y - 1)}
 \end{aligned} \tag{56.2}$$

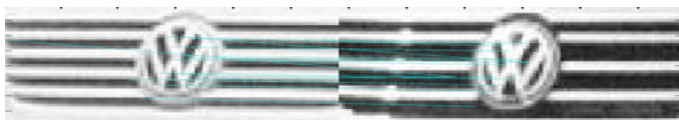
### 56.3.3.2 Feature Matching

In order to match two vehicle face images, the feature matching step is accomplished. For each feature in the query image, the descriptor is used to search for its nearest neighbor matches among all the features from the target image which is going to match. The nearest neighbors are selected by satisfying a minimum Euclidean distance threshold criterion for the descriptor vectors for the query and target image, respectively.

Through the experiments, the matching threshold value in this paper is set to 0.6 (if  $d_{min} < d_{smin} * \text{threshold}$ , success matching). After feature matching, the number of matching feature points is indicated by  $n_3$ . The more the number of matching feature points, the more similar of the two images. The result of feature matching is presented in Fig. 56.4.

### 56.3.3.3 Calculating of the Matching Degree

Combining with the detection (see Fig. 56.3) and matching (see Fig. 56.4) of feature points and between two images, a variable  $n$  is proposed to describe the degree of similarity between images more objectively (Formula 56.3).



**Fig. 56.4** Feature matching

$$n = \frac{1}{2} \left( \frac{n_3}{n_1} + \frac{n_3}{n_2} \right). \quad (56.3)$$

### 56.3.4 Feature Combination

Each feature extracted from the vehicle face by the above three methods characterizes different aspect of image content. The joint exploitation of the different descriptions may provide more comprehensive information in order for a classification system with higher accuracy.

Aiming at vehicle images which needed to be compared, the method in this paper extracts multiple features, combining them to constitute feature vectors for vehicle discriminant and classification. This paper mainly studies three groups of feature vectors, they are color difference  $c$  and vehicle face pattern difference  $m$  (CM[ $c, m$ ]), color difference  $c$  and vehicle logo matching degree  $n$  (CN[ $c, n$ ]), color difference  $c$ , vehicle face pattern difference  $m$  and vehicle logo matching degree  $n$  (CMN[ $c, m, n$ ]).

## 56.4 Experimental Part

### 56.4.1 Data Set and Experimental Process

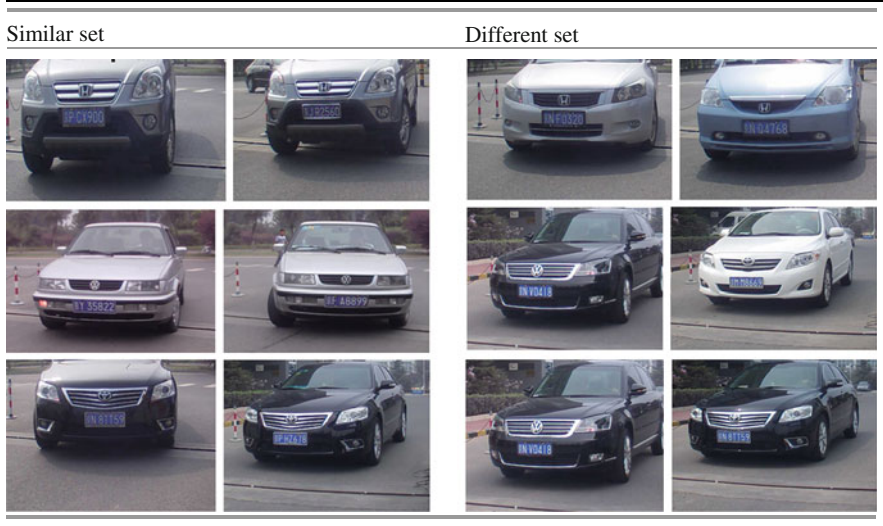
The experimental platform is built on Matlab2010 in this paper and the vehicle data set is constructed using 200 vehicle images collected by parking lot. The proposed method is assessed on a set of mated frontal-view vehicle images from five manufacturers, Volkswagen, Honda, Nissan, Modern and Toyota. Some are similar in the features extracted from vehicle face while others are not. The similar vehicles are divided into similar set and others are as different set by human.

This paper randomly selects 200 pairs of vehicles as the training set (a half from the similar set and others are from the different set), and 100 pairs of vehicles as a test set (a half from the similar set and others are from the different set). As shown in Table 56.1, the vehicles are recognized to similar just when they have the same feature in the color, vehicle face pattern and manufacturer.

In order to assess the method of feature combination more objectively and roundly, this paper compares several feature combination schemes with simple feature scheme, at the same time tries to use a variety of classic classifier to predict the contrast results.

Three groups of feature vectors is chose to experiment, they are color difference  $c$  and vehicle face pattern difference  $m$  (CM[ $c, m$ ]), color difference  $c$  and vehicle logo matching degree  $n$  (CN[ $c, n$ ]), color difference  $c$ , vehicle face pattern

**Table 56.1** Data set



difference  $m$  and vehicle logo matching degree  $n$  ( $CMN[c, m, n]$ ). They are combined together to form the corresponding training set and test set, then they are respectively adopted with decision tree, KNN ( $N = 1$ ) and the Naive Bayes algorithm to operate classification training and test. The expect result of classification is the pair of vehicles are similar or not.

### 56.4.2 Analysis of the Results

The approach of three kinds of feature combination is compared in this paper. As shown in Table 56.2, the results on training set indicates using multiple feature combined method can effectively improve the performance of the vehicle discrimination, especially at the time when the color, the vehicle face pattern and the logo matching degree combined with each other ( $CMN[c, m, n]$ ). The test results on training set indicate that no matter which classifier and test method is used to

**Table 56.2** Test results on training set

Feature combination	CM([c, m])		CN([c, n])		CMN([c, m, n])	
	10-fold cross	66 % training	10-fold cross	66 % training	10-fold cross	66 % training
Decision tree	81	78	80	80	84	82
KNN	63	73	87	64	97	91
Naive Bayes	80	82	90	91	93	91



**Table 56.3** Test results on test set

Feature combination	CM( $[c, m]$ )		CN( $[c, n]$ )		CMN( $[c, m, n]$ )	
	Similar set	Different set	Similar set	Different set	Similar set	Different set
Decision tree	76	78	81	90	91	90
KNN	68	71	80	88	90	92
Naive Bayes	74	80	84	92	92	96
Average	75		86		92	

classification, the feature description method of CMN always show a better performance in accurate rate.

Table 56.3 shows that CMN method can achieve better performance than CM and CN method in the three different classifiers. Its average classification accuracy is 92 %, which is better than the CM performance 75 % and CN performance 86 %. The results could prove the combined feature method of CMN( $[c, m, n]$ ) has higher accurate rate and stronger adaptability. It is worth noting that, the test on different set exhibit better performance than similar set.

## 56.5 Conclusion

The contribution of this paper mainly lies in the effective use of different features extracted in vehicle face to describe the similarity or difference of two vehicles, increasing the possibilities for correct recognition and make the recognition process more simple and robust in practical application. It picks out the optimum combination of features, namely CMN ([color difference, the level of vehicle face pattern difference, vehicle logo matching degree]). Using different classifier to test, and reaching a 94 % classification accuracy on average, could meet the real-time application requirements.

In later work, vehicle face feature can be refined, such as fetching vehicle face texture feature value or a combination of vehicle-logo recognition technology, a more comprehensive description of the vehicle face information and can be improved. Therefore, to improve the classification accuracy and efficiency is a further research focus.

## References

1. Kato T, Ninomiya Y, Masaki I (2002) Preceding vehicle recognition based on learning from sample images. *IEEE Trans Intell Transp Syst* 3(4):252–260
2. Yoshida T, Mohottala S, Kagesawa M, Ikeuchi K (2002) Vehicle classification systems with local-feature based algorithm using CG model images. *IEICE Trans E00-A(12)*:121–126
3. Sivaraman S, Trivedi M (2010) General active-learning framework for on-road vehicle recognition and tracking. *IEEE Trans Intell Transp Syst* 11(2):267–276

4. Xiong N, He J, Park JH, Cooley D, Li Y (2009) A neural network based vehicle classification system for pervasive smart road security. *Univ Comput Sci* 15(5):1119–1142
5. Frintrop S, Rome E, Christensen H (2010) Computational visual attention systems and their cognitive foundations: a survey. *ACM Trans Appl Percept* 7(11):1–46
6. Wang W, Shang Y, Guo J, Qian Z (2011) Real-time vehicle classification based on eigenface. In: *Proceedings of international conference on consumer electronics, communications and networks (CECNet)*, IEEE, pp 4292–4295
7. Zhang B, Zhao C (2011) Classification of vehicle make by combined features and random subspace ensemble. In: *Proceedings of 6th international conference on image and graphics*, pp 920–925
8. Yang M (2011) Vehicle color recognition using monocular camera. In: *Proceedings of wireless communications and signal processing (WCSP)*, pp 978–982
9. Shan Y, Sawhney HS (2006) Shapeme histogram projection and matching for partial object recognition. *IEEE Trans Pattern Anal Mach Intell* 28(4):568–577
10. Petrovic VS, Cootes TF (2004) Analysis of features for rigid structure vehicle type recognition. In: *Proceedings of British machine vision conference*, vol 2, pp 587–596
11. Psyllos AP, Anagnostopoulos C-NE, Kayafas E (2010) Vehicle logo recognition using a SIFT-based enhanced matching scheme. *IEEE Trans Intell Transp Syst* 11(2):322–328

# Chapter 57

## Fast SIFT Algorithm Using Recursive Gaussian Filters

Zhengyuan Ye, Shouxun Liu and Xuan Wang

**Abstract** Scale invariant feature transform (SIFT) algorithm has drawn great attention from computer vision engineers since it was proposed in 1999. However, the high computational complexity of the algorithm has hindered its application. In this paper, a fast SIFT algorithm is proposed, in which FIR Gaussian filters are replaced by recursive filters. Experimental results show that the proposed fast SIFT method needs less computation and yields nearly the same performance compares to original method. It is also recognized that the impulse response approximation error can be used as a good measure to estimate performance degradation of SIFT algorithm in recursive Gaussian filters. Furthermore, through using recursive filters, more choices of the values of prior smoothing scale can be made without considering the number of operations.

**Keywords** SIFT · Recursive Gaussian · Feature points · Object recognition

### 57.1 Introduction

SIFT algorithm, proposed by Lowe [1, 2], is a widely adopted and researched image matching method based on feature points detection and description. Owing to its distinctive performance, SIFT becomes widely used in many computer vision

---

Z. Ye (✉)

School of Information Engineering, Communication University of China,  
No 1 Dingfuzhuang East Street, Chaoyang District, Beijing, China  
e-mail: yzycuc@sina.cn

S. Liu · X. Wang

Engineering center of Digital Audio and Video, Communication University of China,  
No 1 Dingfuzhuang East Street, Chaoyang District, Beijing, China  
e-mail: sxliu@cuc.edu.cn

X. Wang

e-mail: wangxuan@cuc.edu.cn

applications, such as recognition of objects, object tracking and images stitching [3]. Since most computer vision applications demand for real-time performance, the high computation complexity of SIFT algorithm has restricted its usage. One of the most time consuming part of the SIFT is a feature detector, named as DOG detector, in which a number of Gaussian filtering are carried out.

Traditionally an image is used to convolve with FIR Gaussian of different sizes separated by a constant multiplicative factor to produce a family of images at different scale [4]. Then, difference of filtered images is computed at two nearby scales to produce a DOG image which is required by DOG detector. But this process needs a large amount of computations. For 2D images, a convolution kernel of size  $n \times n$  the cost is  $n \times n$  multiplications per pixel by directly convolution. For a separable kernel, as the Gaussian is, this number can be reduced to  $2n$ . This is too high especially when the scale,  $\sigma$ , of the Gaussian is large. For a FIR Gaussian of scale  $\sigma$ , it is customary to use a kernel of size at least  $8\sigma$ . This burden could be too heavy for a real-time performance [5]. Lots of work has been done, like the PCA-SIFT [6], SURF [7] methods to reduce the cost of time. There is also some research has been done in hardware for accelerating [8, 9]. However, no one considers changing the implement of Gaussian filters in SIFT which can make it faster to our knowledge. Due to the cost of the FIR Gaussian filter, lots of work has been done to make Gaussian filtering fast. Among the methods, the recursive filtering approximation to Gaussian is outstanding for its higher accuracy, less operations per pixel and constant operations per pixel independent of  $\sigma$  [10–15].

A modified SIFT method using recursive filtering approximation to Gaussian filtering for real-time system was proposed in this paper. By using recursive filters, the amount of computation in a SIFT algorithm is reduced by at least a factor of about four comparing to FIR filters. The reduction factor is determined by the amount of the prior smoothing chosen in practical application. There is negligible loss in matching performance is registered through our intensive tests. Moreover, being one of noting merits of recursive Gaussian filters, any distend of prior smoothing can be selected without considering the cost of time. This is desirable for large size images or images with abundant details, considering that these images will lead to large number of feature points and a drop in the quality of feature points detected.

In this paper, the content is as follows. In Sect. 57.2.1, the recursive Gaussian filters are introduced, then the filters are extended to SIFT for real time system in our software for tests, our assessment standard mainly depend on the filters' performance and extra benefits compared to FIR Gaussian filter in Sect. 57.2.2 and computations in Sect. 57.3. Conclusion is given in Sect. 57.4.

## 57.2 Performance of Recursive Filtering in SIFT

### 57.2.1 Recursive Filtering

There are three main recursive methods referred by Sovira Tan [5]. These three techniques were implemented using quite different methods through space domain, z domain, Fourier domain respectively. Tan pointed out that both Deriche [12] and Vliet et al.'s filters [15] appear to be convincing candidates to approximate Gaussian convolution at high speed [5]. Therefore tests are focused on Deriche and Vliet et al.'s filters.

For more details about these two recursive filters, readers can refer to the related references. We do not spread them out, due to the reason of paper length.

### 57.2.2 Performance

When recursive Gaussian filters are chosen to be implemented in SIFT algorithm instead of FIR filters, the performances of the two different schemes are measured in repeatability rate of feature points, which is proposed by Cordelia Schmid [16], as well as matching performance of feature points. Also, the impulse responses of different filters are compared as a reference.

#### 57.2.2.1 Impulse Response

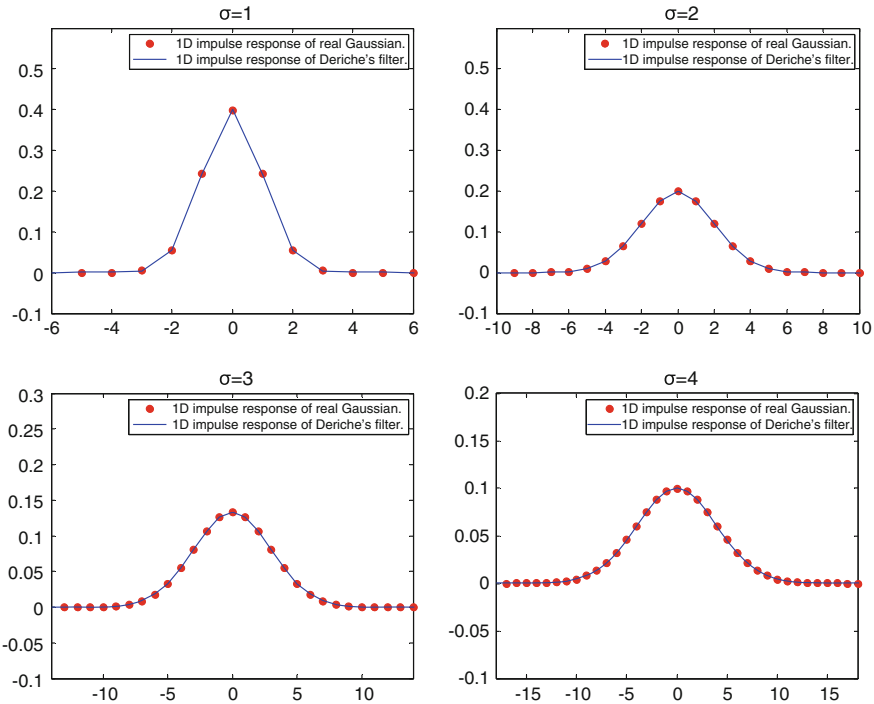
Figures 57.1 and 57.2 show the impulse responses of the recursive Gaussian compare to real Gaussian they designed to approximate.

In the figures, the real Gaussian response is plotted in red circles. It can be seen from these figures that Deriche's filter fits the Gaussian well at all scales. As for Vliet et al.'s filter, some mismatch appears at low scales, but its results become better when the scale increases. It deserved to be mentioned that Tan [5] measured the filters by using a normalized root-mean-square (rms) error measure defined as:

$$\Delta_1 = \frac{1}{N_D} \sum_{(i,j) \in D} \frac{\sqrt{(h(i,j) - g_\sigma(i,j))^2}}{g_\sigma(i,j)} \quad (57.1)$$

This  $h(i,j)$  is the impulse response of the recursive filter in 2D and  $g_\sigma(i,j)$  is the response of real Gaussian.  $D$  means the domain where the distance is shorter than  $3\sigma$  from the peak of the Gaussian.  $N_D$  is the total number of points in  $D$ .

The conclusion of his paper showed that Deriche and Vliet et al.'s filters perform better with increasing scale. The error is below 1 %, as  $\sigma$  increases, it becomes much lower. That suggests that using a recursive filter in the place of a costly FIR Gaussian convolution can be a practical solution [5].

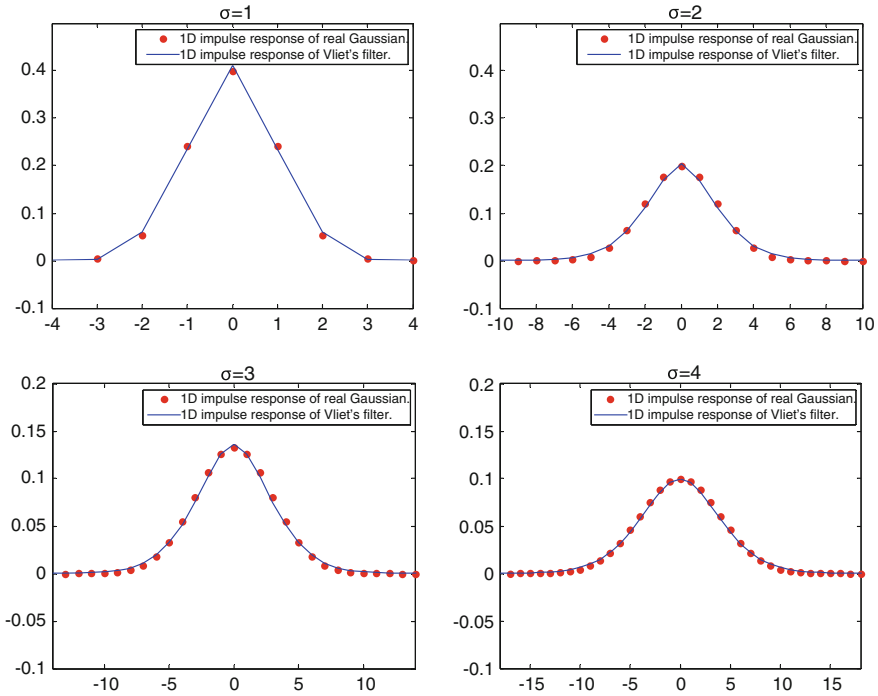


**Fig. 57.1** 1D impulse responses of Deriche's filter against real Gaussian at four scales

### 57.2.2.2 Repeatability

To measure the filters' performance in SIFT, the traditional FIR Gaussian filter module in SIFT is replaced with the recursive Gaussian module. After that, the repeatability rate of feature points is calculated through experiments proposed by Cordelia Schmid [16]. These figures are plotted through experiments using a collection of 20 images from a diverse range. Each image was subject to a range of transformations, including rotation, scaling, affine transformation, little change in brightness and contrast, and addition of image noise. Because the changes were synthetic, it was possible to precisely predict where each feature in an original image should appear in the transformed image, allowing for measurement of correct repeatability for each feature [1].

Figure 57.3 shows that the repeatability continues to increase with  $\sigma$  independent of the choice of the filters. This conclusion is the same as Lowe's, at the same time, Lowe considered using a large  $\sigma$  may lead to less efficiency due to his filtering method, so he had chosen to use  $\sigma = 1.6$ , which provides close to optimal repeatability [1]. However, higher repeatability rate can be obtained results from benefits of the recursive filters. Since the amount of calculations of the algorithm using recursive Gaussian filters is independent of  $\sigma$  (see Sect. 57.1), higher



**Fig. 57.2** 1D impulse responses of Vliet et al.’s filter against real Gaussian at four scales

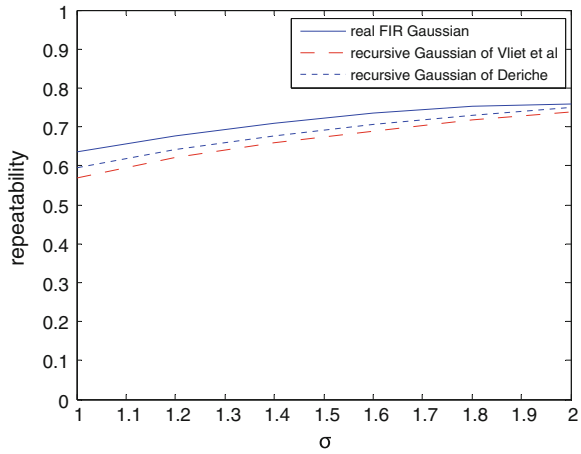
repeatability rate can be obtained through increasing value of  $\sigma$  without cost of more computations. Even so, a problem should be paid attention to which is as prior smoothing  $\sigma$  goes up, the number of detected feature points decreases. And easy to understand, these missing points are the less stable ones than the existing ones. Therefore, as the amount of prior smoothing becomes larger, the detected feature points are more convincing. This is useful especially when less but more stable features are wanted to achieve real time performance in applications.

### 57.2.2.3 Matching Performance

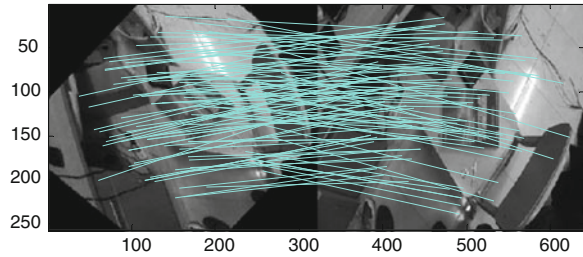
At last matching performance of three filters is tested and one of the samples is shown in Figs. 57.4, 57.5, 57.6. Image in the left side is the transformation of the right side. It can be easily found that the performance of Deriche’s filter and Vliet et al.’s is so close to the real FIR Gaussian filter. In general, the three filters can be used to complete the detection task in many applications.

In fact, other different filters which fit the real Gaussian not so well are used to test. The matching performance is bad due to their impulse responses are less approximation to the real Gaussian than these two filters’. The experimental results are not listed, because the fast algorithm in this paper is independent of them.

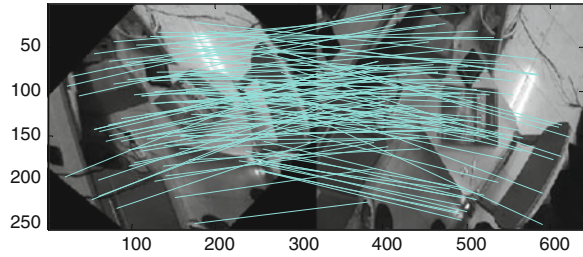
**Fig. 57.3** The repeatability rate of feature points extracted by three different filters. The  $\sigma$  means the amount of prior smoothing, which is applied to each image level before building the scale space representation for an octave



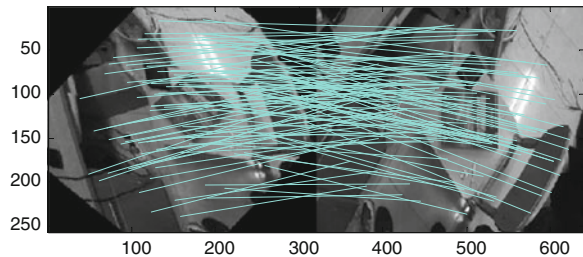
**Fig. 57.4** There are 105 matches at prior smoothing = 1.6 by using FIR Gaussian filter



**Fig. 57.5** There are 103 matches at prior smoothing = 1.6 by using Deriche's filter



**Fig. 57.6** There are 102 matches at prior smoothing = 1.6 by using Vliet et al.'s filter





**Table 57.1** The number of operations for three different filters

Operations	Filter methods		
	FIR Gaussian	Deriche	Vliet et al.
Multiplications	$4 \times [4\sigma] + 2$	24	16
Additions	$4 \times [4\sigma]$	22	12

Through the experiments, an important additional conclusion is drawn that as the impulse responses are more similar, the matching performance is closer.

### 57.3 Computations

In practice, recursive filters can be chosen to use according to our need especially in real time system.

Table 57.1 shows the operations per pixel of the three different filters, where  $[4\sigma]$  means ceil of  $4\sigma$ . The operations of FIR become larger as the  $\sigma$  increases. The computations of the other two are constant which are independent of  $\sigma$ .

### 57.4 Conclusion

At first, the impulse response of the two recursive filters are both good approximation to FIR filters, the influence is related to the similarity of the filters' impulse response. Both Deriche and Vliet et al.'s filters yield enough matches and has a faster speed. For instance, when the FIR Gaussian filter module in our real time tracking application is replaced with the Vliet et al.' filter, the difference of the tracking results are difficult to tell, but to see the time consuming of the filtering reduce greatly at a factor of about four, which is determined by the chosen prior smoothing.

Moreover, both their approximation performance improves with increasing scale, which guides us to take advantage of this characteristic, at the same time, the larger the prior smoothing scale is, the more stable feature points would be got in SIFT. The best of both worlds can be made by increasing the prior smoothing scale at appropriate level by using the recursive filters which can ensure enough stable feature points while eliminating the points with low quality. As the feature points become convincing and less, the feature descriptor and matching work become easier and faster.

Finally, as for recursive filters, changing the scale simply requires changing the filter's coefficients and the structure of the filtering is fixed which is easier to optimize both in software and hardware.

## References

1. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2):91–110
2. Lowe DG (1999) Object recognition from local scale-invariant features. In: *International conference on computer vision—ICCV*, vol 2, pp 1150–1157
3. Faraj A, Danijela R, Axel G (2010) VF-SIFT: very fast sift feature matching. In: *DAGM Symposium for Pattern Recognition*, pp 222–231
4. Lindeberg T (1994) *Scale-space theory in computer vision*. Kluwer Academic Publishers, Dordrecht
5. Sovira T, Jason LD, Alan J (2003) Performance of three recursive algorithms for fast space-variant Gaussian filtering. *Real-Time Imaging* 9(3):215–228
6. Yan K, Rahul S (2004) PCA-SIFT: a more distinctive representation for local image descriptors. *Comput Vis Pattern Recogn* 2:506–513
7. Herbert B, Andreas E, Tinne T, Luc JVG (2008) Speeded-up robust features (SURF). *Comput Vis Image Underst* 110(3):346–359
8. Feng-Cheng H, Shi-Yu H, Ji-Wei K, Yung-Chang C (2012) High-performance SIFT hardware accelerator for real-time image feature extraction. *IEEE Trans Circuits Syst Video Technol* 22(3):340–351
9. Jingbang Q, Tianci H, Takeshi I (2009) A 7-round parallel hardware-saving accelerator for Gaussian and DoG pyramid construction part of SIFT. In: *Asian conference on computer vision—ACCV*, pp 75–84
10. Deriche R (1990) Fast algorithms for low-level vision. *IEEE Trans Pattern Anal Mach Intell* 12(1):78–87
11. Deriche R (1992) Recursively implementing the Gaussian and its derivatives. In: *Proceedings of the second international conference on image processing, Singapore*, pp 263–267
12. Deriche R (1993) Recursively implementing the Gaussian and its derivatives. Research report 1893, INRIA, France
13. Young IT, van Vliet LJ (1995) Recursive implementation of the Gaussian filter. *Signal Process* 44:139–151
14. Jin JS, Gao Y (1997) Recursive implementation of LoG filtering. *Real-Time Imaging* 3(1):59–65
15. van Vliet LJ, Young IT, Verbeek PW (1998) Recursive Gaussian derivative filters. In: *Proceedings of the 14th international conference on pattern recognition, Brisbane, Australia*, pp 509–514
16. Cordelia S, Roger M, Christian B (2000) Evaluation of interest point detectors. *Int J Comput Vis* 37(2):151–172

# Chapter 58

## A Robust Linear Camera Calibration Based on Coplanar Circles

Yu Cai and Yanjin Huang

**Abstract** In this paper, a novel method for camera calibration based on coplanar circles is proposed. Firstly, we describe geometrical interpretation of the proposed method. Then the corresponding strict algebraic proof is presented. Experiment shows that the result is accurate and the proposed strategy is robust.

**Keywords** Camera calibration · Circular points · Coplanar circles · Projective invariant

### 58.1 Introduction

Camera Calibration is a basic problem in computer vision. In order to overcome the shortcoming of the traditional calibration methods which need the special high-precision three-dimensional calibration block [1]. Zhang [2] presented a novel calibration method using planar template instead of the three-dimensional calibration block, the method required locating the physical coordinates of the feature points on the planar template, and required the one-to-one correspondences between template and real image. The cost became large if the number of the features was too large.

In order to get rid of the inconvenience, [3] presented a calibration method based on circular points, the main advantage of this method is to avoid the feature points corresponding step, however, the robustness was poor. Considering a conic needed more parameters than determining a point or a straight line, therefore, the methods based on conics should have a higher robustness. Penna [4] used the conics as the template, and then solved the homography matrix using of the corresponding points. Wright et al. [5] took advantage of the fact that the four

---

Y. Cai (✉) · Y. Huang

Institute of Mathematics, Jilin University, NO 2699, Qianjin Road, Changchun, China  
e-mail: caiyu09@mails.jlu.edu.cn

common tangents of two conics remained unchanged under projective transformation. Hu and Tan [6] used the circle with two intersection points as the template, by solving the intersections of the quadric curves to obtain the images of the circular points. Then solved the camera parameter matrix. Hu et al.[7] could get the camera parameters by the two circles at arbitrary distribution. All of these methods needed to solve the binary quadratic nonlinear equations, although the problem can be transformed into a quartic equation.

In addition, [8, 9] gave the linear calibration algorithms based on the conics, however, these method needed to know the shape parameters and relative position parameters of the conics, also needed to match the conics between template and real image. Presented the linear calibration algorithms based on circles using the curve beam [10, 11]. But the above two methods had a specific requirement of circular distributions. Wherein [10] required two concentric circles, [11] required two or more circles without intersections.

In this paper, we present a linear calibration method using three (or more) circles with arbitrary distributions. It could handle various distributions such as intersect, tangency, separation, etc. The method does not require the geometric information and distribution information of circles and match information between template and image. The calibration process is linear.

The paper is organized in four sections and conclusion. In Sect. 58.2, we present a briefly introduction about some basic concepts and theories involved. In Sect. 58.3, we focus on the principle of the algorithm. In Sect. 58.4 we present the experimental results and data analysis. Finally, conclusions are outlined in Sect. 58.5.

## 58.2 Background Knowledge

### 58.2.1 Camera Imaging Principle

Assuming that the projective transformation between template plane  $[X \ Y]$  and image plane  $[u \ v]$  is:

$$\lambda(u, v, 1)^T = K[r_1 \ r_2 \ t](X, Y, 1)^T.$$

where  $\lambda$  is nonzero scale factor,  $K = \begin{pmatrix} f_u & s & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$  is the camera intrinsic parameter matrix,  $[r_1 \ r_2 \ t]$  is the camera extrinsic parameter matrix,  $r_1, r_2$  are the first two cols of 3D rotation matrix  $R = [r_1 \ r_2 \ r_3]$ ,  $t$  is 3D translation vector. Assuming  $H = K[r_1 \ r_2 \ t]$ , then  $H$  can be seen as homography matrix of projective transformation, and must satisfy  $\det(H) \neq 0$ .

## 58.2.2 Absolute Conics and Circular Points

In a 2D projective space, a point  $p = (x, y, t)^T$  ( $t = 0$ ) is called point at infinity. All of such points form the line at infinity  $l_\infty$ . Its homogeneous coordinates is  $(0, 0, 1)^T$ . Likewise, in a 3D projective space, all points at infinity form the plane at infinity  $\pi_\infty$ , the homogeneous coordinates is  $(0, 0, 0, 1)^T$ . In the plane of infinity  $\pi_\infty$ , points satisfying  $P^T P = 0$  constitute the absolute conic  $\Omega_\infty$ . We can easily verify that the image of  $\Omega_\infty$  is the conic  $K^{-T} K^{-1}$ . This indicates that the image of  $\Omega_\infty$  contains all the information about camera's intrinsic parameters. Hence if we can solve the image of  $\Omega_\infty$ , we can derive the intrinsic parameter matrix.

Without loss of generality, we can assume the model plane lies on the  $X$ - $Y$  plane in the world coordinate system, so the equation of the model plane is  $Z = 0$ . Thus,  $(X, Y, 0, 0)^T$  is intersecting line of plane at infinity and model plane. That is the line at infinity  $l_\infty$  of model plane. In particular, the point  $(1, i, 0, 0)^T$  and  $(1, -i, 0, 0)^T$  (called circular points) of absolute conic lied on line at infinity  $l_\infty$  of model plane. Obviously, the image of conic lied on conic  $K^{-T} K^{-1}$ . Therefore, if we found enough (at least three pairs) images of circular points, it is easy to get camera intrinsic parameters.

Assume  $O$  is arbitrary circle on the model plane. The point on  $O$  satisfies  $(\frac{x}{r} - O_x)^2 + (\frac{y}{r} - O_y)^2 = r^2$ . The intersection points  $(1, i, 0, 0)^T$  and  $(1, -i, 0, 0)^T$  between  $O$  and line at infinity are the circular points [9]. Thus, if we found a circle on the model plane and the image of one line at infinity (generally called the vanishing line), their intersection points are the images of two circular points.

## 58.3 Algorithm Principle

We will introduce the solving principle of the vanishing line from geometry aspect and algebra aspect.

### 58.3.1 Geometric Description

Firstly, Proposition 1 gives the invariance of the correspondence between pole and polar in the projective geometry.

**Proposition 1** *The correspondence between pole and polar remains the same under the projective transformation.*

The proof is simple, omitted here.

**Definition 1** Given one point  $p$  and two circles  $O_1, O_2$ , if  $p$  satisfies: (1) Coplanar with  $O_1, O_2$ . (2) The two polar between  $p$  and  $O_1, O_2$  coincide. Then we call the point  $p$  as the comment pole of  $O_1, O_2$ .

Next we discuss the various possibilities of comment pole. Without loss of generality, we assume the center of the circle  $O_1$  located at coordinate origin, and the radius of circle  $O_1$  is 1. The center of circle  $O_2$  located at  $(d, 0)$ , and the radius is  $r(r \neq 1)$ .

### 58.3.1.1 Concentric Circles ( $d = 0$ )

The two circles' matrix presentments are:

$$O_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, O_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -r^2 \end{pmatrix}$$

Assume  $(x, y, 1)^T$  is the comment pole of  $O_1, O_2$ , and  $(x, y, 1)^T$  is not the point at infinity. Then  $O_1 \cdot (x, y, 1)^T = \lambda \cdot O_2 \cdot (x, y, 1)^T$ , where  $\lambda$  is nonzero scale factor. We can get the following equations:

$$\begin{cases} \frac{x}{-1} = \frac{x}{-r^2} \\ \frac{y}{-1} = \frac{y}{-r^2} \end{cases} \quad (58.1)$$

As the two circles are not coincident, that is  $r^2 \neq 1$ . The Eq. (58.1) only have zero solution. In other words, the comment center is the comment pole of two circles.

Assume  $(x, y, 0)^T$  is the comment pole of  $O_1, O_2$ , and  $(x, y, 0)^T$  is the point at infinity. Satisfying  $O_1 \cdot (x, y, 0)^T = \lambda \cdot O_2 \cdot (x, y, 0)^T$ . Obviously, the equation holds for arbitrary  $x$  and  $y$ . This means that all of points on the line at infinity are the comment pole of two circles.

Through the above analysis, the comment poles of two concentric circles consist of the comment center and all of points on the line at infinity. By Proposition 1, under the projective transformation, the comment poles of images of two circles are exactly the images of center and the vanishing line. In the following sections, we will present the determination condition of the concentric circles and the method for finding the vanishing line.

### 58.3.1.2 Non-concentric Circles ( $d \neq 0$ )

The two circles' matrix presentments are:

$$O_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, O_2 = \begin{pmatrix} 1 & 0 & -d \\ 0 & 1 & 0 \\ -d & 0 & d^2 - r^2 \end{pmatrix}.$$

Assume  $(x, y, 1)^T$  is the comment pole of  $O_1, O_2$ , and  $(x, y, 1)^T$  is not the point at infinity. Then  $O_1 \cdot (x, y, 1)^T = \lambda \cdot O_2 \cdot (x, y, 1)^T$ , where  $\lambda$  is nonzero scale factor. We can get the following equations:

$$\begin{cases} dx^2 + (r^2 - d^2 - 1)x + d = 0 \\ y = 0 \end{cases} \quad (58.2)$$

Since  $d \neq 0$ , the Eq. (58.2) have a pair of conjugate solutions. The necessary and sufficient condition for the two solutions coincide is  $r^2 = (d \pm 1)^2$ . This means that when the two circles are tangent, the comment poles of the two conjugate solutions coincide.

Assume  $(x, y, 0)^T$  is the comment pole of  $O_1, O_2$ , and  $(x, y, 0)^T$  is the point at infinity. Satisfying  $O_1 \cdot (x, y, 0)^T = \lambda \cdot O_2 \cdot (x, y, 0)^T$ . Obviously, the equation holds if  $x = 0$ . This means that the point at infinity  $d_\infty = (0, 1, 0)^T$  is the comment pole of two circles.

Through the above analysis, the comment pole of two non-concentric circles consist of a pair of conjugate points lied on the center line of the two circles and a point at infinity. In the following sections, we will give the method for searching the image  $d_{m\infty}$  of the point at infinity  $d_\infty$  using the comment pole.

### 58.3.2 Algebraic Description

A point  $X = (x, y, t)^T$  lied on the projective plane same as the circles  $O_1, O_2$ . The necessary and sufficient condition for that  $X$  is the comment pole of two circles is  $O_2 X = \lambda O_1 X$ . It can be rewritten as

$$(O_2 - \lambda O_1)X = 0 \quad (58.3)$$

where  $\lambda$  is nonzero scale factor. The images of two circles  $O_1, O_2$  are the conics as  $C_1 = \lambda_1 H^{-T} O_1 H^{-1}$ ,  $C_2 = \lambda_2 H^{-T} O_2 H^{-1}$ . Where  $\lambda_1, \lambda_2$  is nonzero scale factor. We combine the equations and (58.3):  $(C_2 - \lambda \frac{\lambda_2}{\lambda_1} C_1)HX = 0$ . That is:

$$(C_2 - \lambda' C_1)HX = 0. \quad (58.4)$$

where  $\lambda' = \lambda \frac{\lambda_2}{\lambda_1}$ ,  $HX$  is the comment pole of  $C_1, C_2$ .

Since  $H$  is non-singular, and  $\lambda_1, \lambda_2$  are non-zero, Eq. (58.3) is equivalent to Eq. (58.4). This is the algebraic interpretation of Proposition 1.

**58.3.2.1 The Positional Relationship of the Two Circles and the Calculation of the Point at Infinity**

Here, we analyze the characteristic Eq. (58.3). Suppose  $\det(O_2 - \lambda O_1) =$

$$\begin{vmatrix} 1 - \lambda & 0 & -d \\ 0 & 1 - \lambda & 0 \\ -d & 0 & d^2 - r^2 + \lambda \end{vmatrix} = 0, \text{ we have}$$

$$\lambda^3 + (d^2 - r^2 - 2)\lambda^2 + (1 - d^2 + 2r^2)\lambda - r^2 = 0. \tag{58.5}$$

i. If  $d = 0$ , the solutions of Eq. (58.5) is  $r^2$  and 1, where  $\lambda = 1 \neq r^2$  is the double root.

From the expression of  $O_2 - \lambda O_1$ , we can know that if and only if  $\lambda = 1$ , the rank of  $O_2 - \lambda O_1$  is 1. That means Eq. (58.5) has double root  $\lambda$ .

ii. If  $d \neq 0$ , the solutions of Eq. (58.5) is

$$\lambda^{(2)} = \frac{1}{2} - \frac{1}{2}d^2 + \frac{1}{2}r^2 + \frac{1}{2}\sqrt{1 - 2d^2 - 2r^2 + d^4 - 2d^2r^2 + r^4}$$

$$\lambda^{(3)} = \frac{1}{2} - \frac{1}{2}d^2 + \frac{1}{2}r^2 - \frac{1}{2}\sqrt{1 - 2d^2 - 2r^2 + d^4 - 2d^2r^2 + r^4}$$

The necessary and sufficient condition for that Eq. (58.5) has double root  $\lambda$  is  $\lambda^{(2)} = \lambda^{(3)}$ , that is  $r^2 = (d \pm 1)^2$ . Here  $\lambda = \lambda^{(2)} = \lambda^{(3)} = 1 - d \neq \lambda^{(1)}$ . The rank of  $O_2 - \lambda O_1$  is 2. That means if and only if  $d \neq 0$  and  $r^2 \neq (d \pm 1)^2$ , Eq. (58.5) has no double root. On the other hand, if and only if  $d \neq 0$  and  $r^2 = (d \pm 1)^2$ , Eq. (58.5) has double root  $\lambda$ , and  $\lambda$  make the rank of  $O_2 - \lambda O_1$  be 2.

Based on the above discuss, we can obtain the discrimination method for the positional relationship of the two circles  $O_1, O_2$  using Eq. (58.5), which can be expressed as the following proposition

**Proposition 2** (1) If Eq. (58.5) does not have multiple roots, the two circles are not concentric or tangent. By solving the independent real root  $\lambda$  (not conjugate roots), combine with Eq. (58.3), and the solved eigenvector  $X$  is the point at infinity  $d_\infty$ . (2) If Eq. (58.5) has multiple roots, they will be the double root  $\lambda$ , substitute  $\lambda$  to the matrix  $O_2 - \lambda O_1$ , we have

(a) If the rank of the matrix  $O_2 - \lambda O_1$  is 1, that means the two circles are concentric. Substitute  $\lambda$  to Eq. (58.3), the solved two eigenvectors can be spanned as the line at infinity.

(b) If the rank of the matrix  $O_2 - \lambda O_1$  is 2, that means the two circles are tangent. By solving the independent real root  $\lambda'$ , Substitute  $\lambda'$  to Eq. (58.3), the solved eigenvector is the point at infinity  $d_\infty$ .



### 58.3.2.2 Computing the Vanishing Lines

Since Eq. (58.3) is equivalent to Eq. (58.4), the above results are also satisfied for Eq. (58.4). We have the parallel proposition as proposition 2, here the equation corresponding to Eq. (58.5) is

$$\det(C_2 - \lambda' C_1) = 0 \quad (58.6)$$

**Proposition 3** (1) If Eq. (58.6) does not have multiple roots, the two circles of the conics  $C_2, C_1$  are not concentric, and tangent. By solving the independent real root  $\lambda'$ , combine with Eq. (58.4), the solved eigenvector  $HX$  is the point at infinity  $d_\infty$ . (2) If Eq. (58.6) has multiple roots, they will be the double roots  $\lambda'$ , substitute  $\lambda'$  to the matrix  $C_2 - \lambda' C_1$ , we have

- (a) If the rank of the matrix  $C_2 - \lambda' C_1$  is 1, that means the images of  $C_2, C_1$  are concentric. Substitute  $\lambda'$  to Eq. (58.4), the solved two eigenvectors  $HX$  is the image of the point at infinity  $d_{m\infty}$ .
- (b) If the rank of the matrix  $C_2 - \lambda' C_1$  is 2, that means the images of  $C_2, C_1$  are tangent. By solving the independent real root  $\lambda'$ , Substitute  $\lambda'$  to Eq. (58.4), the solved two eigenvectors  $HX$  is the image of the point at infinity  $d_{m\infty}$ .

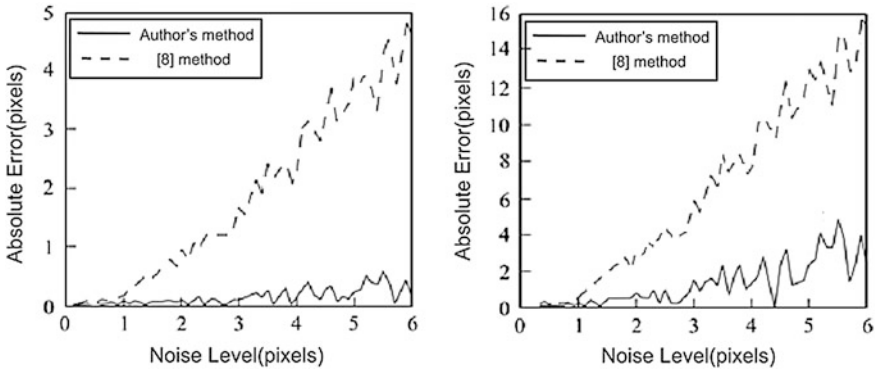
The linear computation is done as [3] did.

## 58.4 Experiment Results

In the simulation experiment, the parameters of the camera are  $f_u = 1200, f_v = 1000, s = 0.2, u_0 = 600, v_0 = 500$ , the resolution of the image is  $1200 \times 1000$ . We use three images, the corresponding position parameters are Axis of rotation:  $n_1 = [0.957841, 0.281718, 0.056343]$   $n_2 = [0.232103, 0.290129, 0.928414]$   $n_3 = [0.198029, 0.693103, 0.693103]$ . Angle of rotation:  $\alpha_1 = \alpha_2 = \alpha_3 = \pi/12$  Translation vector:  $t_1 = [10, 30, 15]^T$ ,  $t_2 = [5, 40, 30]^T$ ,  $t_3 = [10, 30, 30]^T$ , The template includes five coplanar circles (we can extract ten pairs of circles).

Figure 58.1 are the comparative experimental results of the absolute error, including intrinsic parameters  $f_u, u_0$  (since  $f_v$  is similar to  $f_u$ ,  $v_0$  is similar to  $u_0$ , the corresponding results are omitted). From the above comparative results, the proposed method is insensitive to noise, and has higher accuracy and robustness, especially under the calibration condition with high noise.

In order to test the robustness of the proposed method, in this experiment, we test our method using realistic images, we select two different cameras, SONY w120 and Nikon S230, under different resolutions. The resolution of SONY w120



**Fig. 58.1** The comparative results of absolute error of the intrinsic parameters

**Table 58.1** Real calibration results between our method and the other two CLASSIC PLANAR calibration methods

	Parameters	$f_u$	$f_v$	$s$	$u_0$	$v_0$
SONY	Author's method	3502.88	3524.26	-28.73	1110.43	820.91
	[3]	3555.68	3593.43	-38.81	1100.49	786.95
	[2]	3548.91	3613.18	-27.93	1218.28	763.81
Nikon	Author's method	2988.63	3064.80	38.34	1399.35	1080.93
	[3]	3186.13	3324.32	44.43	1388.19	986.59
	[2]	2800.69	3081.25	-29.27	1300.69	995.42
Nikon	Author's method	1848.11	1993.73	-16.37	689.72	381.27
	[3]	1918.45	1999.53	-21.35	703.21	395.25
	[2]	1889.31	2002.63	-19.56	693.31	391.54

is  $2282 \times 1712$ , the resolutions of Nikon S230 are  $2592 \times 1944$  and  $1024 \times 768$ . As we all know, the method proposed in [2] is one of the most popular plane calibration methods, so we also give the result of Zhang's method. The final experimental results are shown in Table 58.1.

### 58.5 Conclusion

In this paper, we propose a simple and efficient camera calibration method, which the solution procedure is linear. Our method only needs three pictures under different angles, and the make of the template is very easy.

## References

1. Tsai R (1986) An efficient and accurate camera calibration technique for 3d machine vision. *Proc Comput Vis Pattern Recognit* 28:364–374
2. Zhang Z (2000) A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell* 22(11):1330–1334
3. Hu ZY, Li H, Meng XQ (2000) A new easy camera calibration technique based on circular points. In: *Proceedings of the British Machine Vision Conference*, pp 496–501
4. Penna MA (1991) Determining camera parameters from the perspective projection of a quadrilateral. *Pattern Recogn* 24(6):533–541
5. Wright J., Wagner A., Shankar Rao, Yi Ma. (2006). Homography from coplanar ellipses with application to forensic blood splatter reconstruction. *Computer Vision and Pattern Recognition*, pp 1250–1257
6. Hu ZZ, Tan Z (2006). Camera calibration with conics fitting and circular points. *J Xi'an Jiaotong Univ* 40(10):1065–1068 (in Chinese)
7. Hu ZY, Li XJ, Wu YH, Wu FC (2006) Coplanar circles, quasi-affine invariance and calibration. *Image Vis Comput* 24(4):319–326
8. Hu ZY, Yang CZ, Sun FM (2000) Planar conic based camera calibration. *Chin J Comput* 23(5):541–547
9. Gurdjos P, Kim JS, Kweon IS (2005) Geometric and algebraic constraints of projected concentric circles and their applications to camera calibration. *IEEE Trans Pattern Anal Mach Intell* 27(4):637–642
10. Hu ZY, Wu FC, Wang GH (2003) A linear approach for determining intrinsic parameters and pose of cameras from rectangles. *J Softw* 14(3):703–712
11. Gurdjos P, Sturm P, Wu Y (2006) Euclidean structure from  $n \geq 2$  parallel circles: theory and algorithms. *ECCV*, pp 238–252

# Chapter 59

## A Multi-Classification Algorithm of Semi-Supervised Support Vector Data Description Based on Pairwise Constraints

Ying Zhao and Guan-jun Wang

**Abstract** A constraint-based semi-supervised support vector machine classification learning algorithm is proposed based on support vector data description algorithm with pairs of semi-supervised learning thinking combined. Multiple hyperspheres are constructed by constraints for the  $k$ -classification problems, so that the original problem converted to a  $k$ -classification problem. The algorithm to get positive constraints label and negative constraints label by calculating the degree  $\mu$ -membership of unlabeled samples, then multiple hyperspheres constructed based on the multi-classification algorithm. Finally, simulation experiments on artificial datasets and UCI datasets to verify the effectiveness of the algorithm.

**Keywords** SVDD · Semi-supervised learning · Multi-classification · Pairwise constraints

### 59.1 Introduction

In many real situations, labeled training data are rather scarce and expensive, while unlabeled data are especially abundant and cheap. Semi-supervised learning tries to exploit unlabeled data to improve learning performance, particularly when there are limited labeled training examples. In the semi-supervised learning the label sample data are less, these data in addition to the clear category labels. Label information—pairwise constraints also has a measure of the relationship between the data sample points. Pairwise constraints can be divided into two kinds of positive constraints and negative constraints are constraints associated data points that belong to the same category, the constraint associated with negative data

---

Y. Zhao (✉) · G. Wang

Department of Computer Science and Technology, China University of Mining and Technology, Xuzhou, 221116, People's Republic of China

e-mail: zhaoying@cumt.edu.cn

points that do not belong to the same class. During the past decade, semi-supervised support vector machine has received significant attention and many approaches have been developed [1–3].

Compared to the information of the labeled sample, constraints are more general, easier access to supervisory information. Semi-supervised learning method based on constraint has achieved success in the semi-supervised dimensionality reduction and semi-supervised classification and semi-supervised clustering application. A Multi-Classification Algorithm of Semi-supervised support vector data description based on pairwise constraints(MS<sup>3</sup>VDD-PC) is proposed, which calculating the degree of membership of the labeled and unlabeled samples to get positive constrain labels and negative constrain labels, labeling the non-target samples, then multiple hypersphere constructed based on the multi-classification algorithm.

Several attempts have been made to solve the non-convex optimization problem associated with S<sup>3</sup>VVM, for example gradient descent [4], semi-definite programming [5], continuation techniques [6], non-differentiable methods, concave-convex procedure [7], branch-and bound algorithms, and deterministic annealing [8].

The rest of this paper is organized as follows. Section 59.2 briefly describes the problem. Section 59.3 presents our MS<sup>3</sup>VDD-PC method. An empirical study is presented in Sect. 59.4. The conclusion and discussion are given in Sect. 59.5.

## 59.2 Problem Descriptions

SVDD algorithm in the training sample data domain description only depends on the sample of the target class. However, in practical applications, not only a sample target class, there are a large number of a sample of non-target classes, these data itself also has a significant data distribution information, together with the training of the two kinds of samples, can be appropriately improved the performance of SVDD.

The multi-classification algorithm of semi-supervised support vector data description based on constraints can be described as follows:

Given an independent and identically distributed training dataset  $G = \{x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_{n+m}\}$ ,  $x_i \in R^d$ ,  $m$  is the number of unlabeled objects and  $n$  is the number of labeled objects with its label  $y_i$ ,  $y_i \in \{1, 2, \dots, k\}$ ,  $k > 2$ . The original problem can be expressed as follows:

$$\begin{cases} \min R^2 + C \sum_i^n \xi_i + C^* \sum_{j=1}^m \xi_j^* \\ s.t. \|\phi(x_i) - a\|^2 \leq R^2 + \xi_i \quad i = 1, \dots, n \\ \|\phi(x_j) - a\|^2 \geq R^2 - \xi_j^* \quad j = 1, \dots, m \\ \xi_i, \xi_j^* \geq 0 \end{cases} \quad (59.1)$$

where,  $a$  is the center of hypersphere,  $R$  is the radius,  $C$  is the trade-off of target objects,  $C^*$  is the trade-off of non-target objects. The Lagrangian of problem (59.1) is expressed by :

$$\begin{aligned}
 L(R, a, \xi, \alpha, \gamma) = & R^2 + C \sum_{i=1}^n \xi_i + C^* \sum_{j=1}^m \xi_j^* - \sum_{i=1}^n \gamma_i \xi_i - \sum_{j=1}^m \gamma_j^* \xi_j^* \\
 & - \sum_{i=1}^n \alpha_i [R^2 + \xi_i - (\phi(x_i) \cdot \phi(x_i) - 2a \cdot \phi(x_i) + a \cdot a)] \\
 & - \sum_{j=1}^m \alpha_j^* [(\phi(x_j) \cdot \phi(x_i) + 2a \cdot \phi(x_j) - a \cdot a) - R^2 + \xi_j^*]
 \end{aligned} \tag{59.2}$$

$\alpha_i, \gamma_i \geq 0, \alpha_j^*, \gamma_j^* \geq 0$ , the dual problem of (59.2) as follows:

$$\begin{aligned}
 \max \quad & \sum_{i=1}^n \alpha_i K(x_i, x_i) - \sum_{j=1}^m \alpha_j^* K(x_j, x_j) - \sum_{i=1}^n \sum_{i=1}^n \alpha_i \alpha_i K(x_j, x_j) \\
 & + 2 \sum_{i=1}^n \sum_{j=1}^m \alpha_i \alpha_j^* K(x_i, x_j) - \sum_{i=1}^n \sum_{p=1}^m \alpha_i^* \alpha_p^* K(x_j, x_p) \\
 \text{s.t.} \quad & \left\{ \begin{array}{l} \sum_{i=1}^n \alpha_i - \sum_{j=1}^m \alpha_j^* = 1 \\ 0 \leq \alpha_i \leq C \\ 0 \leq \alpha_j^* \leq C^* \end{array} \right.
 \end{aligned} \tag{59.3}$$

## 59.3 MS<sup>3</sup>vdd-pc

### 59.3.1 Non-Target Samples Labeled

The labeling of non-target sample is a key issue in the semi-supervised learning. First of all, to facilitate the algorithm description, we give some definition used in the algorithms.

**Definition 1** Absolute Distance  $d_{i,M}$

The distance from the sample  $x_i$  to the center of the sphere  $a^M$  is the *absolute distance*  $x_i$  relative to the hypersphere  $M$ , denoted by  $d_{i,M}$ :

$$d_{i,M} = d(x_i, a^M) = \|x_i - a^M\|$$

The SVDD algorithms exists the following questions:

- (a) SVDD algorithm only considers one class sample trained individually, without taking into account the distribution of other classes. But in the practical application, unavoidable that there are different classes of hypersphere overlap, then the non-target class samples may also fall into multiple hypersphere at the same time.
- (b) SVDD algorithm using the kernel functions, original space is mapped to the unknown high-dimensional spaces; the obtained hypersphere's radius is not the same. However, only the absolute distance cannot fully describe which hypersphere the non-target samples belong to.

Therefore, the possibility of non-target sample belonging to some target class is determined by defining its degree of membership.

**Definition 2** Degree -membership,  $D_{i,M}$

The ratio of  $d_{i,M}$  the absolute distance of sample  $x_i$  and  $R^M$  the square of the radius of hypersphere  $M$ , is the degree -membership of  $x_i$  relative to hypersphere  $M$ , denoted by  $D_{i,M}$ :

$$D_{i,M} = D_{i,M}(x_i, a^M) = \|x_i - a^M\|/R^M \tag{59.4}$$

Formula (59.4) shows that when the sample  $x_i$  is located inside the hypersphere  $D_{i,M} < 1$ ; located on the boundary of the hypersphere  $D_{i,M} = 1$ ; located outside of the hypersphere  $D_{i,M} > 1$ .

For the set of non-target  $G^* = \{x_1^*, \dots, x_m^*\}$ ,  $\forall x_i^* \in G^*$ , degree-memberships  $D_{i,1}, D_{i,2}, \dots, D_{i,k}$  are calculated on all  $k$  hyperspheres, sort of degree-memberships  $D_{i,k_1} \leq D_{i,k_2} \leq \dots \leq D_{i,k_k}$ .

**Definition 3** Positive constraints Label,  $L_{i,P}$

Given non-target sample  $x_i^*$ , the class (or label) which minimum degree-membership corresponding to is the accepted label of  $x_i^*$ , expressed as follow:

$$L_{i,P} = k_1$$

**Definition 4** Negative constraint Label,  $L_{i,N}$

Given non-target sample  $x_i^*$ , the class (or label) which maximum degree-membership corresponding to is the accepted label of  $x_i^*$ , expressed as follow:

$$L_{i,N} = k_k$$

For sample  $x_i^*$ , there is such cases that the degree-memberships are inconsistent, therefore, we need to reset to accept the label and refused to label. Rules are as follows:

If  $D_{i,k_1} = D_{i,k_2}$ , for the target object  $x_i^*$ , the degree-membership to  $k_1$  is equal to the degree -membership to  $k_2$ , reset it to no-target object, set  $L_{i,P} = 0$ .

Similarly, if  $D_{i,k_{k-1}} = D_{i,k_k}$ , set  $L_{i,N} = 0$ .

### 59.3.2 Algorithm Description

The complete details about MS<sup>3</sup>VDD-PC algorithm is described as follow:

Algorithm MS<sup>3</sup>VDD-PC

**Training dataset:**  $G = \{x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_{n+m}\}$ ,  $x_i \in R^d$ ,  $m$  is the number of unlabeled objects and  $n$  is the number of labeled objects with its label  $y_i$ ,  $y_i \in \{1, 2, \dots, k\}$ ,  $k > 2$  ( $x_i, y_i$ )

**Output:** The center of the sphere and the radius of  $k$ -classes hyperspheres

**Initialize:** Train on  $\Omega_k = \{x_1^k, x_2^k, \dots, x_{N_k}^k\}$   $k$  sample subset with SVDD algorithm, get  $k$  hyperspheres  $H_1 \dots H_k$ ,

Step 1: Unlabeled sample marked.

For each unlabeled sample  $x_i^*$ , calculated positive constraints Label  $L_{i,P}$  and Negative constraint Label  $L_{i,N}$ , get accepted label set and rejected label set  $\{(L_{1,P}, L_{1,N}), \dots, (L_{m,P}, L_{m,N})\}$

Step 2: The positive constraints Label set and Negative constraint Label set divided into corresponding subsets  $\Omega_{L_{i,P}}$  and  $\Omega_{L_{i,N}}$ . Training on the mixed sets of  $\Omega_k$ ,  $\Omega_{L_{i,P}}^k$  and  $\Omega_{L_{i,N}}^k$  with S<sup>3</sup>VDD algorithm, get  $k$  new hyperspheres and radius  $\{(a^M, R^M)\}_{M=1}^k$ .

Step 3: Test Degree -membership on  $\Omega_k'$ . If there exist  $\forall x_i$  inconsistent, then reset to accept the label and refused to label, return to step 2, otherwise, end of the algorithm and output the result.

How to test? For any unlabeled example  $x_i$ , compute the Degree -membership  $D_{i,M}$ , get  $L_{i,K^*}$ , then its label is  $K^*$ .

## 59.4 Experiment Results and Discussion

To evaluate the performance of the proposed approach, we have conducted experiments on three UCI datasets (Iria, Glass, Sona) and two artificial datasets(COIL3,COIL 20). These are listed in Table 59.1. Label samples selected at random, the remaining samples as unlabeled samples, in order to avoid the category of the training sample set missing, selecting at least one label samples.

Table 59.2 compared the classification precision attained by “one-to-one”, “ $k$ -SVM”,  $\nabla$ S<sup>3</sup>VM, S<sup>3</sup>VM<sup>light</sup> 10 experiments averaged independently.

In Table 59.2, we can see that MS<sup>3</sup>VDD-PC algorithm relative to the “ $k$ -SVM” and “one-on-one” algorithm in the classification accuracy of all data sets have improved to varying degrees. Algorithm classification accuracy in the category number more COIL20 data set, MS<sup>3</sup>VDD-PC was significantly higher than the “ $k$ -SVM” and “one-to-one” algorithm, the MS<sup>3</sup>VDD-PC algorithm suitable for solving a larger number of multi-category classification problems.



**Table 59.1** Description of experiments data sets

Dataset	Classes	Demension	Train set	Labeled examples
Iria	3	4	150	15
Glass	7	10	17644	70
Sonar	2	60	208	10
COIL3	3	1024	364	30
COIL20	20	1024	1440	100

**Table 59.2** Comparison of classification precision

Dataset	“one-to-one”	k-SVM	$\nabla S^3VM$	$S^3VM^{light}$	$MS^3VDD-PC$
Iria	82.5	80.3	84.3	83.2	87.33
Glass	65.28	61.02	70.67	72.2	75.12
Sonar	77.15	77.15	78.09	80.2	78.12
COIL3	54.18	50.26	40.8	53.7	63.63
COIL20	51.32	44.14	53.09	53.98	69.12

With the improvement of the number of categories of the sample set, the classification accuracy of the algorithm can be seen the “k-SVM” gradually smaller than the accuracy of the classification of the “one-to-one” algorithm, which is mainly due to the increase of the number of categories so that the training sample data in uneven distribution the phenomenon intensified, resulting in a decline in classification accuracy.

Classification accuracy on the multi-classification problem of semi-supervised learning algorithm— $\nabla S^3VM$ ,  $S^3VM^{light}$ —two algorithms presence of unstable, classification accuracy was significantly higher than that in the data set on a Glass, and Iria data set of supervised learning methods; However, other data sets on the classification accuracy is improved is not obvious, and even the classification accuracy still exist in the data set COIL3 decline.

$MS^3VDD-PC$  algorithm compared with  $\nabla S^3VM$  and  $S^3VM^{light}$  algorithms on all data sets were improved classification accuracy. In all the training samples, Sonar is a data set with 2 classes, all the  $MS^3VDD-PC$ ,  $k-SVM$ , “one-to-one”,  $\nabla S^3VM$   $S^3VM^{light}$  have degenerated into a two classification algorithms.

The  $MS^3VDD-PC$  algorithm compared with  $k-SVM$ , “one-to-one”,  $\nabla S^3VM$  three algorithms on two classification problems classification accuracy is also improved, mainly due  $MS^3VDD-PC$  in the process of training in taking into account the non-target sample of non-target samples (two classification problems, including the non-class samples with unlabeled samples) contribution to the classifier. This has to a certain extent, reflects the advantages of semi-supervised learning.

Fig. 59.1 compared the training time by the algorithm  $MS^3VDD-PC$ , “k-SVM”, “one-to-one” on multi-class data sets with the highest classification accuracy when training time (in seconds)

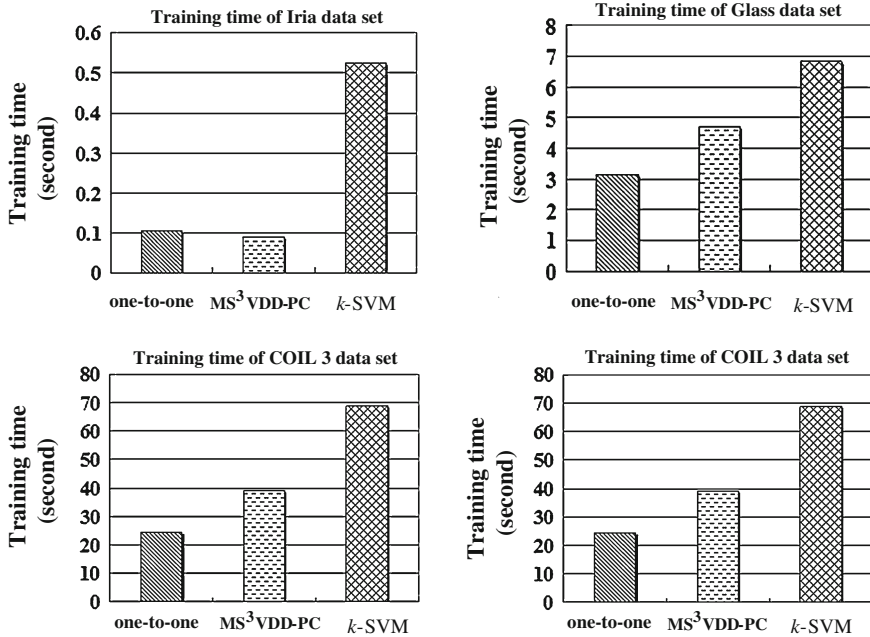


Fig. 59.1 Comparison of training time

In Fig. 59.1, we can see that MS<sup>3</sup>VDD-PC algorithm training time between “one to one” and the *k*-SVM algorithm, the main reason is that MS<sup>3</sup>VDD-PC algorithm in the training process of the class of non-target sample marked, result in an increase in training time. The “one-to-one” algorithm is relatively small due to the number of support vectors generated in the process of training, so training time is relatively short. “*k*-SVM” algorithm improved after a better classification accuracy, but due to the complexity of its objective function, and thus a relatively long training time of the algorithm.

### 59.5 Conclusion

This paper focuses on the multi-classification problem on the characteristics of the support vector data description method, and applied to solve the multi-classification problem and propose a classification learning algorithms based on support vector data domain constraints described. Algorithm is determined by the definition of the membership of the non-target sample positive for label constraints with negative constraints label, constructed on a sample set of non-target in the target sample set and marked hypersphere. The simulation results show that the classification accuracy of this paper MS<sup>3</sup>VDD-PC algorithm in most datasets has improved the actual data set.

**Acknowledgments** This research was supported by the Fundamental Research Funds for the Central Universities (2010QNA28) and the research funds of CUMT (2009).

## References

1. Li YF, Kwok JT, Zhou ZH (2010) Cost-sensitive semi-supervised support vector machine. In: Proceedings of 24th AAAI conference on artificial intelligence, pp 500–505
2. Qi Z, Tian Y, Shi Y (2012) Laplacian twin support vector machine for semi-supervised classification. *Neural Netw* 35:46–53
3. Chakraborty S (2011) Bayesian semi-supervised learning with support vector machine. *Stat Methodol* 8(1):68–82
4. Chapelle O, Zien A, Cowell R et al (2005) Semi-supervised classification by low density separation. *Encycl Biostat* 34:57–64P
5. Bie TD, Cristianini N (2006) Semi-supervised learning using semi-definite programming. In: Chapelle O, Schoëlkopf B, Zen A (eds) *Semi-supervised learning*. MIT Press, Cambridge, pp 119–135
6. Chapelle O, Chi M, Zien A (2006) A continuation method for semi-supervised SVMs. In: Proceedings of international conference on machine learning, Pittsburgh, pp 184–192
7. Collobert R, Sinz F, Weston J, Bottou L (2006) Large scale transductive SVMs. *J Mach Learn Res* 7:1687–1712
8. Sindhwani V, Keerthi v, Chapelle O (2006) Deterministic annealing for semi-supervised kernel machines. In: Proceedings of international conference on machine learning, Pittsburgh, pp 108–116

# Chapter 60

## Real-Time Vehicle Classification Based on Frequency Domain Energy Spectrum

Pengfei Zhang, Haijian Li, Honghui Dong, Limin Jia  
and Maojing Jin

**Abstract** Vehicle classification is now an important part of Intelligent Transportation Systems (ITS). Especially in toll station and parking, real-time vehicle classification technology is used to determine the vehicle information. A novel method based on frequency domain energy spectrum of geomagnetic sensor for real-time vehicle classification was proposed in this paper. According to the definitions of eight frequency domain energy formulations, the energy values with different frequency regions could be computed. Compared with those energy values, the optimal frequency region and energy formulation were obtained. As each vehicle classification has a specific energy region, the classification of each vehicle can be easily differentiated by its energy value. Results show that the vehicle classification method proposed in this paper has an excellent performance and the average accuracy is more than 90 %. Besides, the algorithm makes it easier for applications in sensor nodes with limited computational capability and energy source.

**Keywords** Vehicle classification · Intelligent transportation system · Geomagnetic sensor · Frequency domain · Energy spectrum

---

P. Zhang · H. Li · H. Dong (✉) · L. Jia  
State Key Laboratory of Rail Traffic Control and Safety,  
Beijing Jiaotong University, No 3 ShangYuan Road,  
Beijing, HaiDian District, China  
e-mail: hhdong@bjtu.edu.cn

M. Jin  
High Technology Research & Development Center, The Ministry of Science  
and Technology, Beijing, China  
e-mail: jin@htrdc.com

## 60.1 Introduction

The increasing traffic congestion is a growing problem in many countries. Besides building new roads and bridges to ease congestion, Intelligent Transportation Systems (ITS) seek to maximize the capacity of existing traffic networks and minimize the associated delay [1]. An important function of ITS is information collection, which is performed by traffic flow detectors [2]. Vehicle classification, as an important part of traffic information acquisition, need to be identified in many places such as toll station and parking.

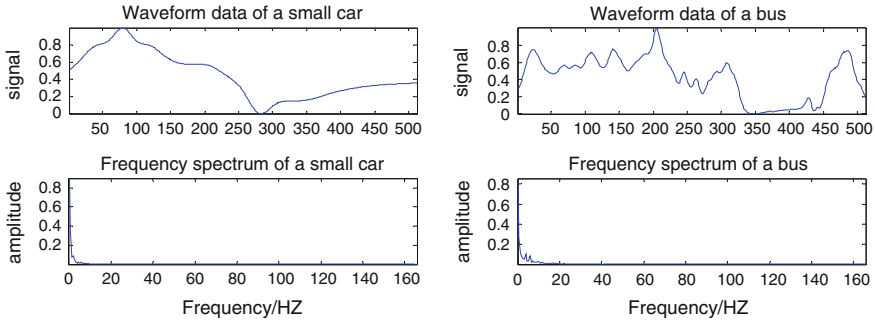
Currently, there are some methods of vehicle classification, such as video image processing [3] and geomagnetic detection [4]. As a new type of vehicle detection equipment, geomagnetic sensor has many advantages, which is easy to install, small footprint and difficult to damage. Many scholars used BP neural network [3] and support vector machine theories [5, 6] to study vehicle classification of geomagnetic sensor, but the drawbacks are the large amount of computation and the uncertain accuracy [6].

In this paper, we proposed a method based on frequency domain energy spectrum to solve the problem of vehicle classification by processing the data collected by geomagnetic sensors. By taking advantage of this method, the process of vehicle classification has a lower amount of computation and the result will be a higher accuracy.

## 60.2 Waveform Standardization

Vehicle data collection is based a multifunctional traffic magnetic sensor designed in our previous study [7, 8]. For all vehicles have significant amounts of iron metals, the magnetic field disturbance created by a vehicle can provide the traffic information detected by a magnetic sensor, which makes it convenient to perform vehicle detection [8]. The magnetic signal outputs from sensors cannot be used directly. The output waveform data of magnetic signal always have some noises, which will influence vehicle detection. It is necessary to process the field waveform data of magnetic signal. In order to facilitate the vehicle classification, the waveform data should be processed by standardization.

Waveform standardization can eliminate the influence to the results of vehicle classification due to different vehicle speeds, and can also avoid the influence of the differences of base magnetic intensity at different time and places. In order to facilitate the comparison of waveform about different vehicle classifications, we just change the proportion of the waveform without changing the shape of the waveform. In the process of waveform standardization, the range of y-axis is set to [0, 1], while the x-axis range is [1, 512]. Figure 60.1 shows the waveforms of a bus and a small car by standardization and their frequency spectrums.



**Fig. 60.1** Waveforms of a small car and a bus and their frequency spectrums

In order to find a binary clustering method to reduce computational cost, we adopt frequency domain to research the method of vehicle classification. We use the Fast Fourier Transform (FFT) to achieve the conversion of frequency domain (see Fig. 60.1). It can be drawn from Fig. 60.1 that only using the feature of frequency spectrum cannot complete vehicle classification. Therefore, further research is needed.

## 60.3 Specific Frequency Domain Energy Spectrums and Vehicle Classification

### 60.3.1 Signal Energy and Vehicle Classification

Following the comparison of the frequency domain, as a characteristic sampling signal, the total energy of a vehicle waveform data are studied. A vehicle waveform which is a limited energy signal is defined as a finite time series. We make use of the energy expression of a time series, and the energy formula of a continuous signal is mentioned as below:

$$E = \int_{-\infty}^{+\infty} |s(t)|^2 dt \quad (60.1)$$

Similarly, the energy formula of a discrete signal is given:

$$E = \sum_{-\infty}^{+\infty} |x(n)|^2 \quad (60.2)$$

After the comparison of the total energy of different vehicle waveforms, we found that the total energy cannot be the basis for vehicle classification. The total energy covers too much content, and the differences of those vehicle waveforms

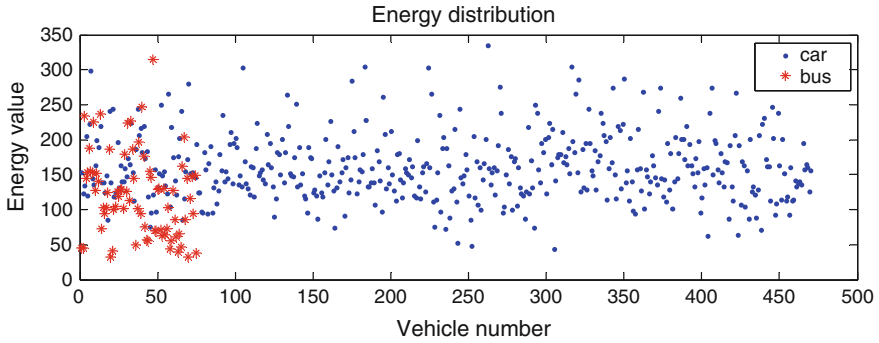


Fig. 60.2 Distributions of total energy of 470 small cars and 75 buses

cannot be obtained in time and frequency domain (Fig. 60.2). In the following study, we observe from different frequency domains. Meanwhile, the energy distribution within a certain frequency range will be studied in detail. Figure 60.3 gives the distributions of frequency spectrum of small cars and buses in some specific frequency domains.

### 60.3.2 Energy Spectrum of Specific Frequency Domain Region

For the energy spectrum of frequency domain, we define eight energy expressions, which are the bases to continue the research of vehicle classification. Since each vehicle waveform data generated as a discrete time serial, eight energy formulations are defined as follows:

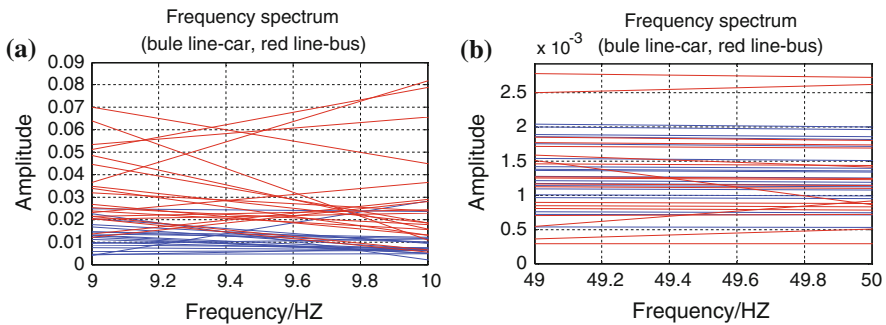


Fig. 60.3 Distributions of frequency spectrum in some specific frequency domains (a 9–10 Hz, b 49–50 Hz)

$$E_1 = \text{sum}f = \sum_0^m f \quad (60.3)$$

$$E_2 = \text{sum}2f = \sum_0^m f^2 \quad (60.4)$$

$$E_3 = \text{sum}Ff = \sum_0^m Ff \quad (60.5)$$

$$E_4 = \text{sum}2Ff = \sum_0^m Ff^2 \quad (60.6)$$

$$E_5 = \text{sum}nf = \sum_0^m nf \quad (60.7)$$

$$E_6 = \text{sum}2nf = \sum_0^m nf^2 \quad (60.8)$$

$$E_7 = \text{sum}Fnf = \sum_0^m Fnf \quad (60.9)$$

$$E_8 = \text{sum}2Fnf = \sum_0^m Fnf^2 \quad (60.10)$$

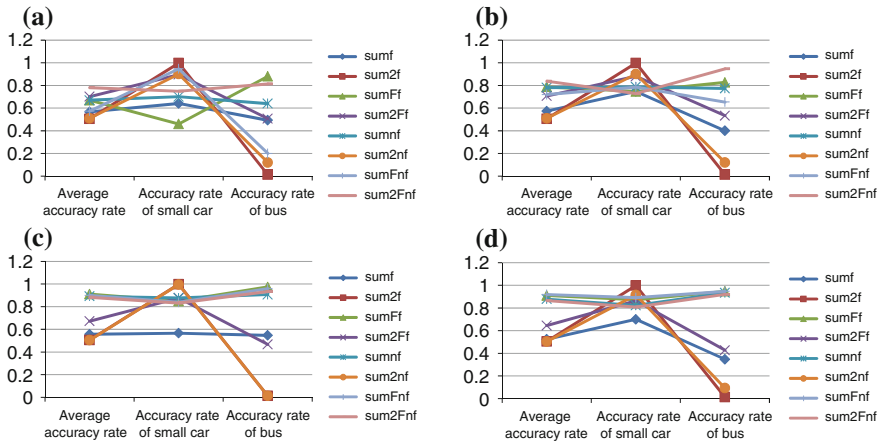
where  $f$  is the amplitude value vector at different frequency, for  $i = 1, \dots, m$ ,  $f(i) \geq 0$ ,  $F$  represents the frequency vector corresponding to the amplitude value vector  $f$ , as a kind of weight of energy,  $n$  represents the number index vector corresponding to the amplitude value vector  $f$ , as another kind of weight of energy,  $m$  is the total length of the amplitude value vector  $f$  for a vehicle between some specific frequency domain region. By the definitions of the energy formula, some results of vehicle classification are shown in [Sect. 60.4](#).

## 60.4 Examples

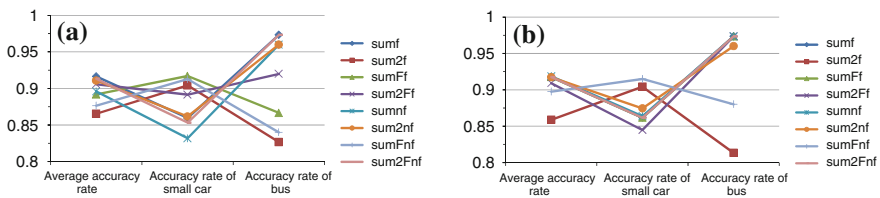
The distributions of each kind of energy in different frequency domain will be studied firstly. When the frequency is greater than 100 Hz, the influence of amplitude of different vehicles to vehicle classification is so weak that could be ignored. Thus, we take the frequency range of 1–100 Hz for analysis.

Figure 60.4 shows the best classification of the energy range in different frequency domain, and it gives the results of the average accuracy rate, the accuracy rate of small car, the accuracy rate of bus in different cases of accuracy rate.





**Fig. 60.4** Best vehicle classification with eight kinds of energy formulation in different frequency domain region (a 1–100 Hz, b 1–5 Hz, c 1–20 Hz, d 1–15 Hz)



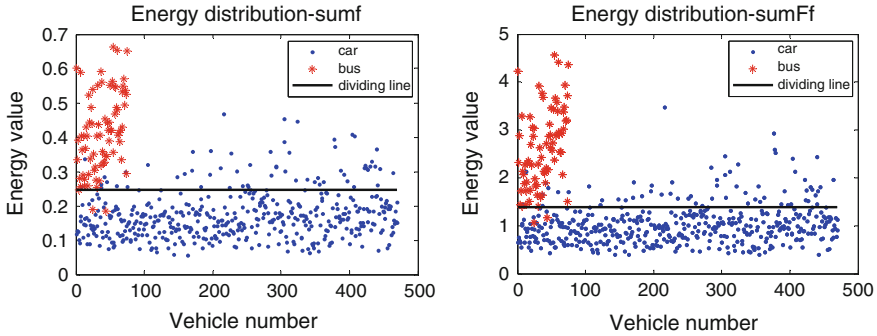
**Fig. 60.5** Accuracy rates of eight kinds of energy formulation in some specific frequency domain regions (a 6–25 Hz, b 6–20 Hz)

Compared with case a, the accuracy rate of small car are overall improved in case b; different from cases a and b, the accuracy rates of average value, small car and bus are all improved in cases c and d, while c is superior than d. At this point, the range of 1–20 Hz is an ideal frequency domain.

After analysis of the frequency spectrum of small cars and buses, it is found that the frequency amplitude spectra of small cars and buses mutually interweave in the range of 1–5 Hz and greater than 25 Hz (as in Fig. 60.3), which will be difficult to distinguish the classifications of different vehicles. And in the range of 6–20 Hz or 6–25 Hz, the frequency amplitude spectra of small cars and buses are significantly separated.

In Fig. 60.5, the accuracy rates of eight kinds of energy formulation in different frequency domain are drawn. By the contrast of Fig. 60.5a and b, accuracy rates of vehicle classification are higher in the range of 6–20 Hz. The best vehicle classification is in the range of 6–20 Hz and specific data is shown in Table 60.1.

For the results of vehicle classification, average accuracy rates in the range of 6–20 Hz are up to 91.85 % (sumnf) or 91.75 % (sumf, sumFf or sum2Fnf). Relatively ideal energy formulations are: E1 (sumf), E3 (sumFf), E5 (sumnf) and E8



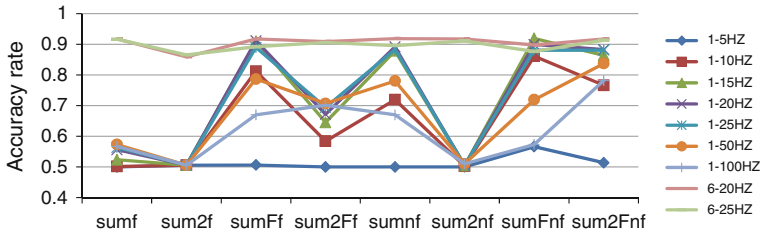
**Fig. 60.6** Vehicle classification results of sumf (the optimal dividing energy value is 0.246) and sumFf (the optimal dividing energy value is 1.392)

**Table 60.1** Accuracy rates of eight kinds of energy formulation in 6–20 Hz

Formulation	Dividing line	Average rate	Rate of small car	Rate of bus
sumf	0.245885544	0.91751773	0.861702128	0.973333333
sum2f	0.010289586	0.85879433	0.904255319	0.813333333
sumFf	1.391826338	0.91751773	0.861702128	0.973333333
sum2Ff	0.031626399	0.90900709	0.844680851	0.973333333
sumnf	2.398333247	0.91858156	0.863829787	0.973333333
sum2nf	0.062790235	0.91723404	0.874468085	0.96
sumFnf	18.70471698	0.89744681	0.914893617	0.88
sum2Fnf	0.298346627	0.91751773	0.861702128	0.973333333

(sum2Fnf). The accuracy rates of bus of these four energy formulations are high to 97.33 %, which is satisfactory. And the accuracy rates of small car are just above 86 %, which is to be improved. As can be seen in Fig. 60.6, the small cars and buses can be basically distinguished in the study, the optimal dividing lines are also been given, which will makes it easier for the application in a sensor node with limited computational capability and energy source.

Finally, we have a comprehensive evaluation of eight energy formulations in different frequency domain regions. Overall, four better energy formulations are: sumFf, sumnf, sumFnf and sum2Fnf. From Fig. 60.7, the accuracy rates of energy formulations within each frequency domain have an explicit show. In the present study, 6–20Hz and 6–25Hz are the best cases of the study. For these two cases, the accuracy rates of all the energy formulations are higher than 80 %, which is an ideal accuracy rate for the field application of vehicle classification.



**Fig. 60.7** Performance comparison of eight energy formulations in different frequency domain regions

## 60.5 Conclusions

In this paper, the method of vehicle classification by using the energy spectrum within different frequency domain regions has been proposed. Through the definition of eight kinds of energy spectrum formulations, we can determine the appropriate frequency domain to the research and analyze the accuracy rate of the vehicle classification. The effective energy formulations and best frequency domain are obtained by field vehicle data. The results of the example show that vehicle classification can be completed by the method proposed in this paper and the accuracy rates are acceptable. Besides, the algorithm used in this method is easy to be implemented in a sensor node with limited computational capability and energy source.

**Acknowledgments** This work is supported by the National Natural Science Foundation of China (61104164) and the Fundamental Research Funds for the Central Universities (2012YJS059), and is also supported by the National 863 Program of China (2012AA112401).

## References

- Cheung SY, Varaiya P (2007) Traffic surveillance by wireless sensor networks: final report. Institute of Transportation Studies, University of California, Berkeley 25
- Li R, Jia L (2009) On the layout of fixed urban traffic detectors: an application study. *Intell Transp Syst Mag* 1(2):6–12
- Wang Z, Liu F (2007) Automobile type of intelligent transportation systems based on genetic & BP algorithms. *Ind Control Comput* 20(9):77–79 (in Chinese)
- Su D, Wang L, Ma S (2007) Vehicle detection method based on magnetoresistive sensors. *Comput Commun* 3(25):9–13 (in Chinese)
- Fan Y, Li M, Zhang Y (2008) Research on application of relevance vector machine in car model identification. *Comput Eng Des* 29(6):1510–1515 (in Chinese)
- Li W, Tao H (2010) Study on vehicle type identification algorithm based on least squares support vector machine. *J Highw Transp Res Dev* 27(1):101–105 (in Chinese)
- Jia L, Dong H (2009) The new traffic sensors, sensor networking optimization and data fusion of traffic state: final report. Beijing Jiaotong University State Key Laboratory of Rail Traffic Control and Safety, Beijing (in Chinese)
- Li H, Dong H, Jia L, Xu D, Qin Y (2011). Some practical vehicle speed estimation methods by a single traffic magnetic sensor. In: International IEEE conference on intelligent transportation systems, pp 1566–1573

# Chapter 61

## A New Combination Sampling Method for Imbalanced Data

Hu Li, Peng Zou, Xiang Wang and Rongze Xia

**Abstract** Imbalanced data is commonly in the real world and brings a lot of challenges. In this paper, we propose a combination sampling method which resamples both minority class and majority class. Improved SMOTE (ISMOTE) is used to do over-sampling on minority class, while distance-based under-sampling (DUS) method is used to do under-sampling on majority class. We adjust the sampling times to search for the optimal results while maintain the dataset size unchanged. Experiments on UCI datasets show that the proposed method performs better than using single over-sampling or under-sampling method.

**Keywords** Imbalanced data · Under-sampling · Over-sampling · SMOTE

### 61.1 Introduction

Traditional machine learning methods assume that the sample distribution is uniform, but many practical problems in the real world are not consistent with this hypothesis. Usually, the useful information is only a small part of the entire dataset, such as the classification of medical data, identification of gene encoding information, forecast of thunderstorms, detection of network intrusion and etc. Therefore, it is very necessary to research on imbalanced data and finds the useful small part data, which is so-called minority class. Since the number of samples in minority class is much fewer than in majority class, traditional classifiers tend to classify all the test samples as majority class, which lead to a higher error rate on the minority class. Assume the case that 99 % of the data belongs to the majority class, 1 % of the data belongs to minority class. If the classifier simply classifies all data as the

---

H. Li (✉) · P. Zou · X. Wang · R. Xia  
School of Computer, National University of Defense Technology, Changsha 410073,  
Hunan Province, People's Republic of China  
e-mail: lihu@nudt.edu.cn

majority class, it can get 99 % accuracy rate, but the really useful information contained in minority class is ignored. Therefore, it is a hot issue on how to effectively deal with imbalanced data in today's machine learning and data mining field.

At present, methods used to deal with imbalanced data can be divided into three categories [1]: (1) methods based on data resampling; (2) methods based on modify the classification algorithms and (3) combination methods of the two. The first method focuses on changing the distribution of the imbalance data, so as to reduce the degree of imbalance. The second method tries to modify the traditional classification algorithm according to characteristics of imbalanced data. The last method aims to combine advantages of both aforementioned methods.

This paper is structured as follows. In Sect. 61.2, related works of resampling are discussed. In Sect. 61.3, we first describe the improved SMOTE method and distance-based under-sampling method. Then a combination sampling method is introduced. Experiments are described in Sect. 61.4. Finally, we present the conclusions and acknowledgements.

## 61.2 Related Works

There are a lot of researches on imbalanced data resampling method, and most of which are in the form of over-sampling and under-sampling. The original over-sampling method copy sample randomly to increase the proportion of minority class, and is likely to cause the classifier over-learning. Original under-sampling method removing samples in majority class randomly, and may delete useful samples. Against drawbacks of original methods, many scholars have proposed different improved methods. Tomek [2] proposed to compute distance between different samples at first, as so to find so-called Tomek links. Then, negative samples were deleted from the Tomek links. Chawla [3] proposed the synthetic minority over-sampling technique (SMOTE). The method creates synthetic minority class samples by taking each minority class sample. Gustavo [4] combined the over-sampling and the under-sampling method, proposed SMOTE + Tomek link method. Taeho [5] did over-sampling based on clustering, which can deal with the imbalance between different classes and within each class. Hui [6] presented two new minority over-sampling methods based on SMOTE, named borderline-SMOTE1 and borderline-SMOTE2, in which only the minority examples near the borderline were over-sampled. Ashish [7] first computed the distance between samples and the decision plane, then the Fisher ratio of samples in minority class respected to samples in majority class was computed. At last, under-sampling was done based on the Fisher ratio. Suzan [8] improved SMOTE method, and apply it on imbalanced time series data. Enislay [9] proposed a hybrid sampling method using SMOTE and Rough Set Theory. Fernandez [10] reviewed the main issues of imbalanced problem and concluded that the imbalance ratio was a key factor, but there other issues that must be taken into consideration such as minority class size, class overlapping and dataset shift. Luengo [11] analyzed SMOTE-based oversampling and evolutionary under-

sampling using data complexity measures, in which C4.5 and PART method were used to learn and classify the data.

## 61.3 Combination Sampling Method

In this section, we first describe two single resampling methods, and then a combination method is introduced. The first method is Improved SMOTE (ISMOTE), which over-sample the minority class. The second one is Distance-based Under- Sampling (DUS) method, which under-sample the majority class. At last, we combine the two methods together and adjust the sampling times so as to find the best results. In order to minimize the effect on time and space complexity for training and classifying the dataset, we keep the dataset size unchanged during the resampling process.

### 61.3.1 Improved SMOTE

SMOTE [3] is a classic over-sampling method, which randomly select several adjacent samples for each minority class sample at first, and then create new minority samples along the line connecting the sample and its neighboring samples. SMOTE assign all neighboring samples the same weight when creating new sample. But samples belong to different class have variant effect on the classifier: minority class samples are inclined to correct the decision plane, while majority class samples may distort the decision plane. So, we propose to use the weight vector, and assign neighboring minority class samples a higher weight while neighboring majority class samples a lower weight. Suppose the minority class contains  $N$  samples and  $R$  is the sample times, then the algorithm is as follows:

Step 1: Select a sample  $x_i$  ( $i = 1, \dots, N$ ) of the minority class, and compute distance between  $x_i$  and other samples in the entire dataset. Then, find the  $k$ -nearest neighbor of sample  $x_i$ .

Step 2: Select  $R/N$  samples randomly from the  $k$ -nearest neighbor samples, record as  $y_j$  ( $j = 1, \dots, R/N$ ).

Step 3: Create new minority class sample  $x'_i$  between  $x_i$  and  $y_j$ .  $x'_i = x_i + \text{random}(0, 1) \times (y_j - x_i) \times \text{Weight}[\text{isMinority}(y_j)]$ , where  $\text{random}(0, 1)$  generate a random number between 0 and 1.  $\text{Weight}[]$  is the weight vector which assign different weight to different class sample.  $\text{isMinority}()$  is used to determine whether sample  $y_j$  belongs to minority class or not.

Step 4: Repeat the above steps for the remaining samples in the minority class, until all samples in the minority class are processed.

Step 5: Merge the new created minority class samples and the original data together, generating the new dataset.

### 61.3.2 Distance-Based Under-Sampling Method

During classification process based on the decision plane, e.g. Support Vector Machine (SVM), only these samples nearby the decision plane play the role. Samples far away from the decision plane make little contribution to the classifier, or even deviates the decision plane from the right place, which lead to a decrease in classification accuracy. So, it is feasible to delete those samples far away from the decision plane. We use Euclidean distance as standard and do under-sampling on majority class. Suppose the minority class contains  $N$  samples and the majority class contains  $M$  samples, then the algorithm is as follows:

Step 1: Select a sample  $z_i$  ( $i = 1, \dots, M$ ) of the majority class, compute distance between  $z_i$  and all samples in the minority class, record as  $d_{ij}$  ( $j = 1, \dots, N$ ).

Step 2: Compute the average distance  $A_i$ :  $A_i = \left( \sum_{j=1}^N d_{ij} \right) / N$ .

Step 3: If  $A_i$  is greater than the predefined threshold,  $z_i$  is deleted, otherwise, reserve  $z_i$ .

Step 4: Repeat the above steps for the remaining samples in the majority class, until all samples in majority class are processed.

Step 5: Reserved samples generate the new dataset.

### 61.3.3 Combination Sampling Method

The above two methods resample the dataset separately. Next, we combine the two methods together and resample the dataset simultaneously. In order to keep the size of the dataset unchanged, number of new created minority class samples should equal to the number of majority class samples deleted. In ISMOTE, we adjust the sample times by changing parameter  $R$ , but the DUS cannot precisely control the number of samples to be deleted. So, we modify the DUS when used together with ISMOTE. Suppose the minority class contains  $N$  samples and the majority class contains  $M$  samples, then the modified algorithm is as follows:

Step 1: For all samples  $z_i$  ( $i = 1, \dots, M$ ) in the majority class, compute distance between  $z_i$  and all samples in the minority class, record as  $d_{ij}$  ( $j = 1, \dots, N$ ).

Step 2: Compute the average distance  $A_i$ :  $A_i = \left( \sum_{j=1}^N d_{ij} \right) / N$ , and sort  $A_i$  in descending order.

Step 3: Delete sample  $z_i$  of  $A_i$  in top- $R$ , and reserve the remaining.

Step 4: Reserved samples generate the new dataset.

**Table 61.1** Statistics of the four selected datasets

Dataset	#Samples	#Attributes	Class (min, maj)	#Samples (min/maj)	Proportion (min:maj)
cmc	1473	10	(2, remainder)	333/1140	1:3.4
Glass	214	10	(3, remainder)	17/197	1:11.6
Haberman	306	4	(2, remainder)	81/225	1:2.8
Pima	768	8	(1, 0)	268/500	1:1.87

## 61.4 Experiments

All experiments were done on a PC with Intel (R) Core (TM) i5-3210 M CPU@2.50 G-Hz, 10G memories and 500 G hard disk capacities. We did all experiments on Matlab R2011b. Four imbalanced datasets namely cmc, glass, haberman and pima selected from UCI datasets<sup>1</sup> were used to test our resampling method. Detailed statistical information of the four dataset is shown in Table 61.1. The original UCI dataset contains more than one class, so we first specified the minority class, and then merged all other classes as the majority class.

### 61.4.1 Evaluation Criteria and Test Method

Evaluation criteria for balanced data, such as precision and recall, are not very effective when used on imbalanced data [3]. At present, the widely used evaluation criteria include the area under the Receiver Operating Characteristic curve: AUC [12] and the geometric mean of the accuracies measured separately on each class: g [13]. In addition, SVM is used to classify resampled dataset. Because the original dataset is not divided into a training set and a test set, in this paper, we use cross-validation method. The whole dataset is randomly divided into four parts at first, and then three of them were taken as training set and rest one was taken as test set. We take the average of the four tests as the final result.

### 61.4.2 Results and Analysis

Firstly, we used ISMOTE, DUS method and the combination sampling method to resample the four datasets. Then, SVM was used to classify the resampled datasets. The results are shown in Tables 61.2 and 61.3.

As can be seen from Tables 61.2 and 61.3, ISMOTE performed better than DUS method under AUC evaluation criteria, but performed worse than DUS under

<sup>1</sup> The dataset is available at <http://archive.ics.uci.edu/ml/datasets.html>



**Table 61.2** AUC of different methods

	cmc	Glass	Haberman	Pima
ISMOTE	0.6890	0.9391	0.6909	0.7588
DUS	0.6789	0.9245	0.6706	0.7635
ISMOTE + DUS	0.6945	0.9350	0.6942	0.7721

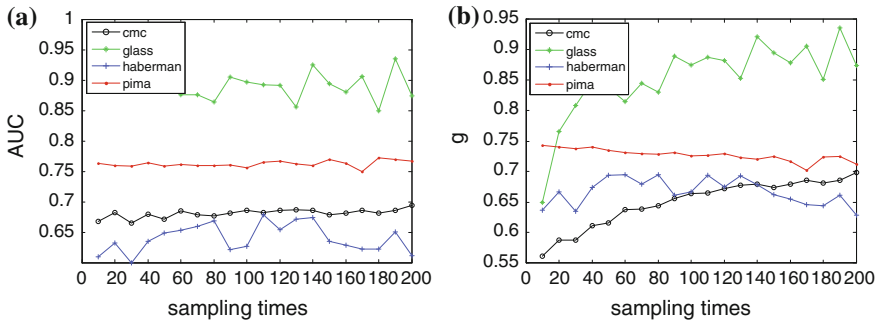
**Table 61.3** g of different methods

	cmc	Glass	Haberman	Pima
ISMOTE	0.6857	0.9327	0.6978	0.7548
DUS	0.8794	0.9220	0.7194	0.7653
ISMOTE + DUS	0.6984	0.9361	0.7134	0.7686

g evaluation criteria. AUC reflect the relation of true positive rate and false positive rate, so the more the minority class samples been classified correctly, the greater of it. ISMOTE created new minority class samples and increased the proportion of minority class, which resulted in the improvement of AUC. Higher g means that the accuracy of both majority class and minority class are high, and the two is balanced which means the difference of the two cannot be too large. DUS did under-sampling on the majority class, and reduced the number of majority class samples, which narrowed the difference between the accuracy of both majority class and minority class. So, DUS performed better under g. The combination method ISMOTE + DUS took the advantages of both above methods and performed best most of the time whether under AUC or g. Therefore, the proposed combination sampling method is effective.

Next, we used the combination sampling method to investigate the trend of AUC and g when sampling times changed. The dataset size was kept the same while the proportion of minority class and majority class within the dataset was changed. The number of new created minority class samples in each test equaled to the number of samples deleted from majority class. The sampling times varied from 10 to 200. The results are shown in Fig. 61.1.

Figure 61.1 shows that the AUC of each dataset kept relative stable when sampling times increased while g of different datasets varied from each other when sampling times changed. The g of glass and cmc went upward gradually, while haberman and pima went downward. Both AUC and g of glass were significantly greater than the other three datasets. As listed in Table 61.1, glass and cmc have higher min/maj proportion than haberman and pima, the proportion of glass even reach at 1:11.6. When the min/maj proportion is relative small, the classifier performs well without resampling. As more and more new minority class samples were created while majority class samples been deleted, the property of the dataset changed gradually, which lead to the decrease of g. On the contrary, when the min/maj proportion is relative great, resampling correct the decision plane, thus improved the performance. Therefore, the proposed method is suitable for the case that the dataset is extremely imbalanced. In addition, we can see that the AUC focuses on the overall



**Fig. 61.1** Trend of AUC (a) and g (b) against different sampling times

classification performance, and does not sensitive to changes within the dataset. Compared with AUC,  $g$  can reflect the subtle changes within the dataset, and thus more conducive to improve sampling methods based on the test results.

## 61.5 Conclusions

There are lots of imbalanced data in our daily lives. Traditional classification methods are not suitable for the imbalanced data, and thus, the study on imbalanced data is of important value. In this paper, we proposed improved SMOTE and distance-based under-sampling method against minority class and majority class. The two methods were combined to search for the optimal result. Experiments on four imbalanced UCI datasets showed that the proposed method performed better than the single use of an under-sampling or over-sampling method. In addition, we found that  $g$  is more sensitive than AUC, and therefore  $g$  is more suitable for use in the case one adjust the sampling strategy according to the classification results.

**Acknowledgments** The authors' work was sponsored by the National High Technology Research and Development Program (863) of China (2010AA012505, 2011AA010702, 2012AA01A401 and 2012AA01A402), the Nature Science Foundation of China (60933005, 91124002), Support Science and Technology Project of China (2012BAH38B04, 2012BAH38B06) and Information Safety Plan (242) of China (2011A010).

## References

1. Haibo H, Edwardo GA (2009) Learning from imbalanced data. *IEEE Trans Knowl Data Eng* 21(9):1263–1284
2. Tomek I (1976) Two modifications of CNN. *IEEE Trans Syst Man Cybern* 6(11):769–772
3. Chawlan V, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 16(1):321–357

4. Gustavo EA, Batista PA, Ronaldo C (2004) A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor* 6(1):20–29
5. Taeho J, Nathalie J (2004) Class imbalances versus small disjuncts. *SIGKDD Explor* 6(1):40–49
6. Han H, Wang W, Mao B (2005) Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: *Advances in intelligent computing, international conference on intelligent computing*, vol 3644, pp 878–887
7. Ashish A, Ganesan P, Fogel GB, Suganthan PN (2010) An approach for classification of highly imbalanced data using weighting and under sampling. *Amino Acids* 39(5):1385–1391
8. Suzan K, Longin JL (2011) Improving SVM classification on imbalanced time series data sets with ghost points. *Knowl Inf Syst* 28(1):1–23
9. Ramentol E, Caballero Y, Bello R, Herrera F (2012) SMOTE-RSB\*: a hybrid preprocessing approach based on oversampling and under sampling for high imbalanced data-sets using SMOTE and rough sets theory. *Knowl Inf Syst* 22(2):1–21
10. Fernandez A, Garcia S, Herrera F (2011) Addressing the classification with imbalanced data: open problems and new challenges on class distribution. In: *Proceeding of the 6th international conference on hybrid artificial intelligent systems* pp 1–10
11. Luengo J, Fernandez A, Garcia S, Herrera F (2011) Addressing data complexity for imbalanced data sets: analysis of SMOTE-based oversampling and evolutionary under sampling. *Soft Comput* 15(10):1909–1936
12. Bradley A (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn* 30(6):1145–1159
13. Kubat M, Matwin S (1997) Addressing the course of imbalanced training sets: one-sided selection. In: *4th international conference on machine learning*. Moran Kaufmann, San Francisco, pp 179–186

# Chapter 62

## Breast Tissue Segmentation Using KFCM Algorithm on MR images

Hong Song, Feifei Sun, Xiangfei Cui, Xiangbin Zhu  
and Qingjie Zhao

**Abstract** Breast MRI segmentation is useful for assisting the clinician to detect suspicious regions. In this paper, an effective approach is proposed for segmenting the breast into different regions, each corresponding to a different tissue. The segmentation work flow comprises three key steps: MR Images preprocessing, locating breast-skin and breast-chest wall boundary by using OTSU thresholding algorithm, and segmenting fibroglandular and fatty tissues with applying the kernel-based fuzzy clustering algorithm (KFCM). The proposed method was applied to segment the clinical breast MR images. Experimental results have been shown visually and achieve reasonable consistency.

**Keywords** Breast MRI · KFCM · Breast tissue segmentation

### 62.1 Introduction

MR images can greatly improve the cure rate of the breast cancer. MR breast images segmentation is to extract different breast regions, each corresponding to a different tissue. It is useful for assisting the clinician to detect suspicious regions, and it can be applied in quantitative analysis of breast tissues, different breast images registration, breast modelling, deformation simulation and assisted surgery guide.

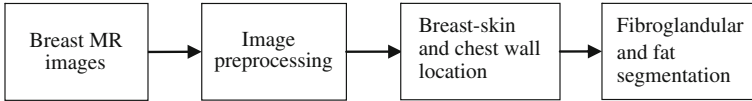
---

H. Song (✉)

School of Software, Beijing Institute of Technology, 5 South Zhongguancun Street,  
Beijing, Haidian District, China  
e-mail: anniesun@bit.edu.cn

F. Sun · X. Cui · X. Zhu · Q. Zhao

School of Computer Science and Technology, Beijing Institute of Technology,  
5 South Zhongguancun Street, Beijing, Haidian District, China



**Fig. 62.1** Breast tissue segmentation framework

Several algorithms have been developed for segment the breast tissues. Threshold-based method, the gradient method, polynomial approximation method, the active contour models and classifier segmentation are used in breast-skin segmentation [1]. Chen et al. [2] applied the fuzzy clustering algorithm to the tumor region segmentation which had achieved better results. Kannan et al. [3] implemented the breast region segmentation by using a new objective function of fuzzy c-means with the help of Hypertangent function, Lagrangian multipliers method, kernel functions. However, these studies did not segment the fat and fibroglandular tissues. Pathmanathan [4] suggested a region-growing method to extract fibroglandular tissue, but it required the user to manually choose one or more seed points, which is inefficient and time consuming. Nie et al. [5] used two steps to segment the breast: firstly, locating the skin boundary and lungs region by use of the standard Fuzzy C-means (FCM) Clustering algorithm, the second step reused an adaptive FCM algorithm to extract the fibroglandular tissue. But this is a semi-automated method.

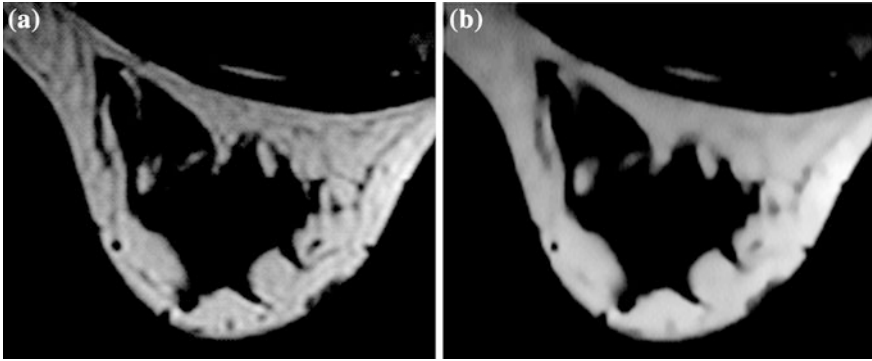
In this paper, we develop an automatic method for a robust segmentation of the breast tissues. The framework of the segmentation method is shown in Fig. 62.1. Firstly, the anisotropic diffusion filter is employed in preprocessing the MR breast images. Secondly, the threshold-based method and morphological operations is used to locate the breast-skin boundary and breast-chest wall. Finally, a kernel-based function fuzzy clustering algorithm (KFCM) is applied to extract the fibroglandular and the fatty tissues. The proposed method was applied to segment the clinical breast MR images. Experimental results have been shown visually and achieve reasonable consistency.

## 62.2 Image Preprocessing

MR image has inhomogeneity, noise and other factors in imaging process which affect the continuity and accuracy of the images segmentation results. Therefore, the anisotropic diffusion filter is used to reduce the image noise. This method was proposed by Perona and Malik [6] which is widely used in image preprocessing. The equation of anisotropic diffusion is defined as:

$$I_t = \text{div}(c(x, y, t)\nabla I) = c(x, y, t)\Delta I + \nabla c \nabla I \quad (62.1)$$

where  $\text{div}(\cdot)$  represents the divergence operator,  $\nabla$  denotes the gradient operator,  $\Delta$  represents the Laplace operator.



**Fig. 62.2** Image preprocessing. **a** Original image and **b** after anisotropic diffusion

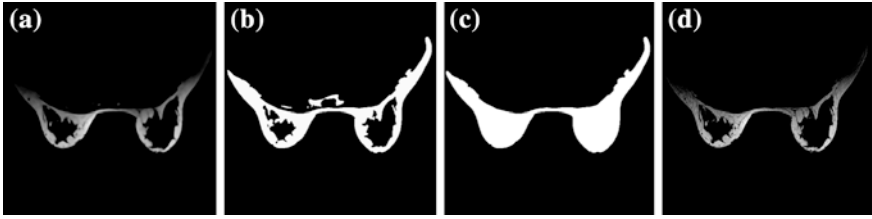
This equation employs a monotonically decreasing function to determine the intensity of diffusion. It's diffuse weak in the region where gradient is large, thereby to ensure that boundary can be better preserved, and intensively in the region where gradient is low, so that the noise can be eliminate and obtain a smooth image while maintaining the edge feature. Figure 62.2a is original breast MR image that is partially enlarged. Figure 62.2b is the image that is processed by anisotropic diffusion filter which is smoother than the original image while the edges are well preserved.

### 62.3 Detecting the Boundary of the Breast Region

Detecting the boundary of the breast region is to locate the breast-skin and breast-chest wall. Since air produces almost zero MR signals, the background in MR images is virtually black, the breast-skin boundary is relatively clear. Pectoral muscle appears dark gray in the axial T1-weighted breast images and the gray level of this area is close to zero, as shown in Fig. 62.3a. So the direct threshold-based method is used to segment the breast region. The method contains the following two steps:

- Step 1: Binarizing the image using OTSU thresholding algorithm

The OTSU thresholding algorithm [7] is applied to the images after preprocessing to separate the breast region from the background. It assumes that the image to be separate contains two classes: foreground and background, then search for the optimum threshold maximizes the inter-class variance. The thresholded image  $g(x, y)$  is defined as:



**Fig. 62.3** Segmentation of the breast region. **a** Result of an anisotropic diffusion, **b** result of OTSU thresholding, **c** after mathematical morphology operations, **d** the breast region

$$g(x, y) = \begin{cases} 1, & f(x, y) \gg T_{OTSU} \\ 0, & f(x, y) < T_{OTSU} \end{cases} \quad (62.2)$$

$T_{OTSU}$  is the threshold computed by the OTSU thresholding algorithm.  $f(x, y)$  denotes the original image. The pixel  $(x, y)$  for which  $f(x, y) \gg T_{OTSU}$  is called an object point (the breast region), otherwise, the point is called a background point. The result of the binarization is shown in Fig. 62.3b.

- Step 2: Mathematical morphology operations

To remove the disturbance of the thoracic spine and the noises, the biggest connected region is selected. The holes of the region are filled with morphology operators. In order to achieve smoothing, we perform a morphological opening followed by a closing. The result of this step is shown in Fig. 62.3c. The extracted breast region is shown in Fig. 62.3d.

## 62.4 Segmentation of the Fat and Fibroglandular Tissue

The interior of the breast region is made up of two major tissue types, fibroglandular tissue and fat. Since the transitional regions between these tissues are not clearly defined, and their memberships are intrinsically vague, in this paper, we used a Kernel-Based Fuzzy C-Means (KFCM) clustering segmentation algorithm which is a non-supervised method to separate the fibroglandular tissue from the fat. In T1-weighted breast MR images, fibroglandular tissue shows low intensity and appears dark gray, fatty tissue shows hypointensity and appears white or light gray, as shown in Fig. 62.3a. This feature is used for FCM-based clustering segmentation.

### 62.4.1 Fuzzy C-Means Clustering

The FCM clustering algorithm is an improvement of the hard k-means algorithm proposed by Bezdek [8]. It has been applied successfully in many areas, e.g. target recognition and images segmentation [9].

The FCM clustering algorithm is used to partition a set of given image pixels into  $c$  clusters, depending on the similarity of the data point to a particular class relative to all other classes, where the number of cluster  $c$  is usually set by a priori knowledge. The standard FCM objective function of partitioning an image into  $c$  clusters is:

$$J_m(u, v) = \sum_{k=1}^n \sum_{i=1}^c u_{ik}^m \|x_k - v_i\|^2 \quad (62.3)$$

subject to:

$$\sum_{i=1}^c \mu_{ik} = 1, 1 \ll k \ll n, \quad \mu_{ik} \in [0, 1], \quad 1 \ll k \ll n, 1 \ll i \ll c \quad (62.4)$$

where  $X = (x_1, x_2, \dots, x_k, \dots, x_n)$  is a data matrix with the size of  $p \times n$ ,  $p$  represents the dimension of feature vector  $x_k$ ,  $n$  represents the number of feature vectors.  $\mu_{ik}$  is the membership of the  $x_k$  vector in the cluster  $c_i$ ,  $v_i$  is the fuzzy cluster centroid of the  $i$ th cluster.  $\|\cdot\|$  is the Euclidean norm, it measures the similarity between a feature vector  $x_k$  and a cluster centroid  $v_i$  in the feature space. The parameter  $m$  is a constant which controls the fuzziness of the resulting partition. We choose  $m = 2$ . The objective function is minimized under the restriction of (62.4).

The main drawbacks of the FCM clustering algorithm are that the algorithm is only suitable for ‘spherical’ clusters and sensitive to noise and outliers. One of the important causes of the non-robust results is the use of Euclidean distance. In this paper, we use a kernel-based FCM clustering algorithm which replaces the Euclidean distance metric by a kernel-induced metric to deal with these problems.

### 62.4.2 Kernel-Based Fuzzy C-Means Clustering

The kernel-based FCM clustering method uses the Mercer’s Theorem. Define a nonlinear map as  $\Phi : (\chi \rightarrow F (x \in R^p \rightarrow \Phi(x) \in R^q, q > p))$ .  $\chi$  denotes the original space of the input data and  $F$  is the transformed feature space with higher dimensions. Then we cluster the data in the transformed feature space.

The KFCM algorithm was proposed by Zhang [10, 11]. This method uses a new kernel-induced metric in the data space to replace the original Euclidean norm metric in FCM. The similarity metric of the (62.3)  $\|x_k - v_i\|$  is replaced by the kernel induced distance  $\|\Phi(x_k) - \Phi(v_i)\|$ . This method has been used successfully in brain image segmentation [10, 11].

The KFCM algorithm minimized the following objective function:

$$J_m(U, v) = \sum_{k=1}^n \sum_{i=1}^c u_{ik}^m \|\Phi(x_k) - \Phi(v_i)\|^2 \quad (62.5)$$



$$\|\Phi(x_k) - \Phi(v_i)\|^2 = K(x_k, x_k) + K(v_i, v_i) - 2K(x_k, v_i) \quad (62.6)$$

We choose the Gaussian Kernel as the kernel function and (62.5) can be written as:

$$J_m(U, v) = \sum_{k=1}^n \sum_{i=1}^c u_{ik}^m (2 - 2K(x_k, v_i)) \quad (62.7)$$

Then we obtain the expressions (62.8) and (62.9) of the membership matrix and cluster centers by the use of LaGrange multiplier method:

$$u_{ik} = \frac{(1 - K(x_k, v_i))^{1/(1-m)}}{\sum_{j=1}^c (1 - K(x_k, v_j))^{1/(1-m)}} \quad (62.8)$$

$$v_i = \frac{\sum_{k=1}^n \mu_{ik}^m K(x_k, v_i) x_k}{\sum_{k=1}^n \mu_{ik}^m K(x_k, v_i)} \quad (62.9)$$

where  $i = 1, 2, \dots, c$ ,  $k = 1, 2, \dots, n$ . Due to the non-convexity of (62.5), the initial value of the membership matrix  $U$  and cluster centers  $V$  has deep influence to the results. Zhang [10] did not give the initializing method, so in this paper we use the results of the FCM algorithm as the input of the KFCM algorithm.

The main steps of the KFCM algorithm can be listed as follows:

Step 1: Initialize the number of clusters  $c$ , the membership matrix  $U$  and  $\varepsilon > 0$  for a very small value as the iteration stopping criterion using the results of the FCM algorithm.

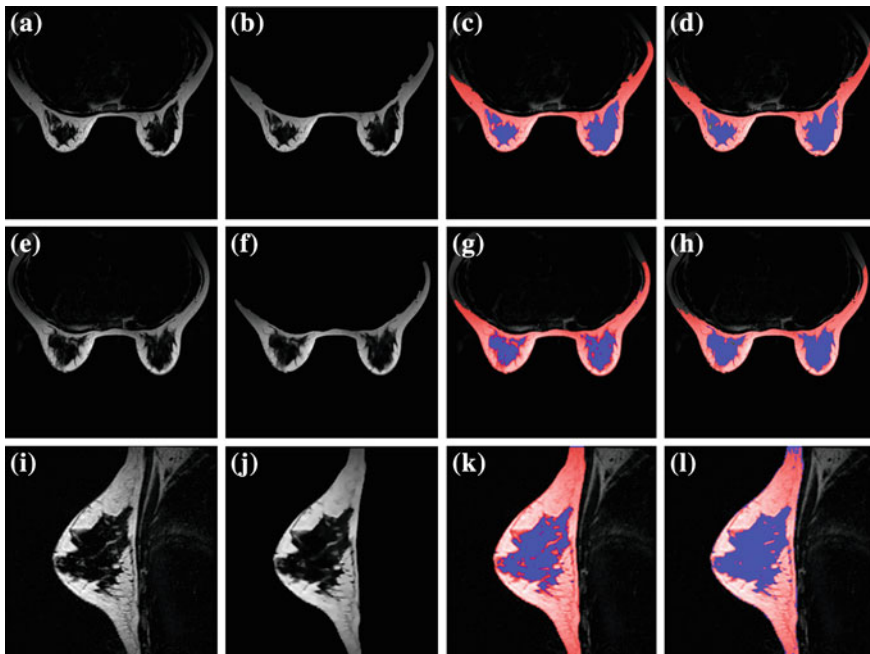
Step 2: Update the membership matrix  $u_{ik}$  (62.8) and the centroids  $v_i$  (62.9).

Step 3: Repeat Step 2 and Step 3 until the criterion  $\|v_{new} - v_{old}\| < \varepsilon$  is satisfied.

## 62.5 Results and Evaluations

The breast MR images we used in this paper were acquired from the Military General Hospital of Beijing PLA. Three data sets were used to evaluate the performance of the proposed algorithm. All procedures were implemented by using MATLAB on a 2.8 GHz CPU, 2G memory PC. All of the breast image slices are T1-weighted. The resolution is  $512 \times 512$  pixels. Figure 62.4a–h are the breast images in transverse orientation. Figure 62.4i–l are the image in sagittal plane. Figure 62.4a, e, i show the original images; Fig. 62.4b, f, j present the breast region extracted from the breast boundary; Fig. 62.4c, j, k show the results of the standard FCM; Fig. 62.4d, h, l show the results of the KFCM. The red regions represent the fat and the blue regions represent the fibroglandular tissue.

We evaluated the segmentation results in terms of some performance indexes in the following: recall, overlap and precision [2].



**Fig. 62.4** The results of the segmentation: original images (a, e, i), the breast region (b, f, j), the results of the standard FCM (c, g, k), the results of the KFCM (d, h, l)

Recall is defined as the number of pixels in the intersection of the automatic segmentation and the manual segmentation, divided by the number of pixels in the manual segmentation alone (62.10). This parameter is an important indicator of the accuracy of segmentation results. Overlap is computed as the number of pixels in the intersection of the automatic segmentation and the manual segmentation, divided by the number of pixels in the union of the automatic segmentation and the manual segmentation (62.11). Precision computed as the number of pixels in the intersection of segmentation and the manual segmentation, divided by the number of pixels in the automatic segmentation alone (62.12).

$$\text{recall} = \frac{A \cap M}{M} \quad (62.10)$$

$$\text{overlap} = \frac{A \cap M}{A \cup M} \quad (62.11)$$

$$\text{precision} = \frac{A \cap M}{A} \quad (62.12)$$

where  $M$  represents the manual segmentation obtained from expert radiologists.  $A$  denotes the result of automatic segmentation.

**Table 62.1** Indexes of the performance

	FCM			KFCM		
	Recall (%)	Overlap (%)	Precision (%)	Recall (%)	Overlap (%)	Precision (%)
Dataset 1	92.04	79.12	79.16	95.00	84.93	83.39
Dataset 2	86.00	73.53	83.53	95.23	79.26	87.76
Dataset 3	85.78	69.90	89.10	95.73	80.80	91.81

The performance indexes of the method are listed in Table 62.1. From this table we can see that the KFCM algorithm is better than traditional FCM algorithm and the parameter recall has been improved above 95 %. It indicates that most of the manually segmented regions are included in the automatic segmentation which could be able to satisfy our requirement. Meanwhile it can be seen by comparing the first two sets of data and the third set that this method is efficient both in transverse and sagittal breast image, which demonstrates the robustness of the proposed algorithm.

## 62.6 Conclusion

In this paper, an automatic method has been proposed for breast tissue segmentation on MR images. We firstly used a threshold-based method to detect the breast-skin and breast-chest wall, so that the breast region can be extracted. Then we used the KFCM algorithm to separate the fat and fibroglandular tissue. The results of segmentation by the KFCM algorithm have been compared with the traditional FCM algorithm. It is obviously that the KFCM performed better than FCM and have reasonable consistency.

**Acknowledgments** Project supported by the National Natural Science Foundation of China (Grant No. 61240010).

## References

1. Raba D, Oliver A, Marti J, Peracaula M, Espunya J (2005) Breast segmentation with pectoral muscle suppression on digital mammograms. In: Proceedings of the 2nd Iberian conference (IbPRIA 2005), Estoril, Portugal. Springer, Berlin
2. Chen W, Giger M, Bick U (2006) A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images. *Acad Radiol* 13(1):63–72
3. Kannan SR, Ramathilagam S, Sathya A (2009) Robust fuzzy c-means in classifying breast tissue regions. In: International conference on advances in recent technologies in communication and computing
4. Pathmanathan P (2006) Predicting tumour location by simulating the deformation of the breast using nonlinear elasticity and the finite element method. Wolfson College University of Oxford, Oxford

5. Nie K, Chen JH, Chan S, Chau MKI, Yu HJ, Bahri S, Tseng T, Nalcioglu O, Su M (2008) Development of a quantitative method for analysis of breast density based on three-dimensional breast MRI. *Med Phys* 35(12):5253–5262
6. Perona P, Malik J (1990) Scale-space and edge detecting using anisotropic diffusion. *IEEE Trans Pattern Anal Mach Intell* 12(7):629–639
7. Otse N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybernet SMC-9*(1):62–66
8. Bezdek JC (1981) *Pattern recognition with fuzzy objective function algorithms*. Plenum Press, New York
9. Wang J, Kong J (2008) A modified FCM algorithm for MRI brain image segmentation using both local and non-local spatial constraints. *Comput Med Imaging Graphic* 32:685–698
10. Zhang DQ (2004) *Kernel-based associative memories, clustering algorithms and their applications*. PhD Dissertation, Nanjing University of Aeronautics and Astronautics (in Chinese)
11. Zhang DQ (2003) Kernel-based fuzzy clustering incorporating spatial constraints for image segmentation. In: *Proceedings of the 2nd international conference on machine learning and cybernetics*, vol 4, pp 2189–2192

# Chapter 63

## Information-Theoretic Clustering for Gaussian Mixture Model via Divergence Factorization

Jiuding Duan and Yan Wang

**Abstract** Multivariate Gaussian distribution is a multi-dimensional generalized model for conventional pervasive Gaussian distribution and Gaussian mixture model (GMM) is a practical model widely used in clustering analysis. In this paper, we aim at clustering the Gaussian mixture components into reduced number of components, thus simplifying the GMM. By utilizing a class of Bregman divergences in information-theoretic context, we minimize the in-cluster relative entropy, and derived a closed-form of optimal parameters by divergence factorization. Symmetric relative entropy is used to avoid the asymmetric manner of divergence. Our algorithm is time-saving as a simplification of GMM and demonstrates its superiority visually in practical clustering-based image segmentation problem, in comparison with conventional methods, such as k-means and GMM-EM.

**Keywords** Clustering · Gaussian mixture · Relative entropy · Bregman divergence

### 63.1 Introduction

Clustering is a fundamental technique for knowledge discovery in various machine learning and data mining applications. Numerous clustering algorithms have been proposed with different practical intuition and theoretical foundation.

In general, we can categorize the clustering algorithms into two classes. One class consists of non-generative approaches such as k-means [1] and spectral clustering algorithm [2]. These methods are designed mainly based on certain heuristics and lack of formulated expressions in statistics thus cannot prove their

---

J. Duan (✉) · Y. Wang

School of Automation Science and Electrical Engineering, Beihang University,  
XueYuan Road No.37, HaiDian District, BeiJing, China  
e-mail: jiuding.duan@gmail.com

efficiency when data is noisy with internal complex structure. Another class is generative approaches such as Gaussian mixture models (GMM-EM) [1]. These methods have some basic model assumptions for data distribution, and the algorithm is designed to solve the optimal model parameters. They provide a general framework for clustering with explicit statistical meaning and ability of incorporating various intuitions.

However, GMM-EM still has some practical problems. First, the computational cost of GMM-EM is not satisfactory for large scale data clustering. Researchers have made progress on simplifying the mixture models to induce more practical algorithms. Zhang and Kwok [3] proposes a linear complexity clustering algorithm by minimizing an upper bound on the approximation error between the original and simplified models. Goldberger et al. [4] uses unscented transform, a method originally introduced for filtering nonlinear dynamical systems, to learn a simplified representation of Gaussian mixture. Nielsen et al. [5] uses KL-divergence to design a k-means like clustering algorithm for Gaussian mixtures.

In this paper, we proposed an information-theoretical clustering algorithm based on divergence factorization to interpret GMM into an optimal assignment for mixtures by divergence measures. It employs a class of Bregman divergence measure [6] to derive the optimal solution, thus achieve an iterative clustering with satisfactory computational cost.

## 63.2 Preliminaries

### 63.2.1 Multivariate Gaussian Mixture

Multivariate Gaussian distribution is an important model that is widely used in many application problems. It is a natural generalization of standard univariate case of naïve Gaussian distribution and often used to describe, at least approximately, any set of correlated real-valued random variables.

The probability density function (PDF) of a d-dimensional multivariate Gaussian distribution can be modeled by

$$p(x|\mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2}\right) \quad (63.1)$$

where  $\mu \in R^d$  is the mean and  $\Sigma > 0$  is the covariance which is a symmetric positive-definite matrix.

Mixture model, especially Gaussian mixture model (GMM) is widely used in various domains to discover potential sub-domain information. The basic idea of GMM is based on the practical assumption that the samples are composed by mixture of finite multivariate Gaussian components. A GMM can be modeled by

$$f(x) = \sum_{i=1}^n \alpha_i f_i(x) \quad (63.2)$$

where  $f_i(x) = p(x|\mu, \Sigma)$  and  $\alpha_i$  is the weight of each component, with  $\sum_{i=1}^n \alpha_i = 1$ .

A typical method on solving parameters is EM algorithm which aims to maximize the likelihood of sampled data in an unsupervised manner [1]. EM algorithm excessively uses the data samples iteratively which increases the computational cost when dealing with large datasets.

### 63.2.2 Symmetric Relative Entropy

Relative entropy is an important concept in information theory. It measures a difference between statistical distributions and is also called Kullback–Leibler divergence (KL divergence). Relative entropy between two distributions is defined as

$$D(f||g) = \int f(x) \log \frac{f(x)}{g(x)} \quad (63.3)$$

where  $f(x) = p(x|\mu, \Sigma)$  and  $g(x) = p(x|m, S)$  are probability density functions for identical distributions. The above form is asymmetric which may lead to unstable measurement. So a symmetric measure should be used to evaluate the ‘distance’, it can be defined as

$$SD(f, g) = D(f||g) + D(g||f) \quad (63.4)$$

Bregman divergence is a class of strictly convex function [6] and is defined as

$$D_\phi(x, y) = \phi(x) - \phi(y) - \langle x - y, \nabla \phi(y) \rangle \quad (63.5)$$

where  $\phi$  is a real-valued, strictly convex and differentiable function defined over a convex set  $Q = \text{dom}(\phi) \subset \mathbb{R}^d$ .  $\nabla \phi(y)$  represent the gradient vector of  $\phi$  evaluated at  $y$ .

For example, Mahalanobis distance  $M_{\Sigma^{-1}}(x, y) = (x - y)^T \Sigma^{-1} (x - y)$  is a widely used Bregman divergence. The corresponding convex function is  $\phi(x) = x^T \Sigma^{-1} x$ , where  $\Sigma^{-1}$  is the inverse of covariance matrix.

Also, Bregman divergence can be extended to matrices

$$D_\phi(X, Y) = \phi(X) - \phi(Y) - \text{tr}((\nabla \phi(Y))^T (X - Y)) \quad (63.6)$$

where  $X$  and  $Y$  are matrices,  $\phi$  is a real-valued, strictly convex and differentiable function defined over matrices, and  $\text{tr}(A)$  is the trace of matrix  $A$ . In specific, Berg matrix divergence is a class of divergence generated from convex function  $\phi(x) = \|X\|_F$ , and can be written as

$$B_\phi(x, y) = \text{tr}(XY^{-1}) - \log|XY^{-1}| - d \quad (63.7)$$

The symmetric relative entropy, Mahalanobis distance and Burg divergence are three important components that enable our clustering algorithm.

### 63.3 Information-Theoretic Clustering via Divergence Factorization

Since divergence is a measure to evaluate the ‘distance’ between distributions [7]. The intuition of our clustering algorithm is minimizing the overall in-cluster divergence. It clusters the original clusters in ways of assigning each original cluster to a specific cluster. This results in re-assigning cluster index for each cluster, while the samples initialized into original clusters remain the same. It can be viewed as an alternative way for simplifying the GMM in ways of reducing number of clusters.

For  $n$  multivariate Gaussians parameterized by mean vector  $m_1, \dots, m_n$  and covariance  $S_1, \dots, S_n$ , the objective is to find an optimal disjoint partitioning of the  $n$  Gaussians into  $k$  clusters,  $\pi_1, \dots, \pi_k$ . Symmetric relative entropy is used to evaluate the ‘distance’ between distributions, and the overall objective function can be described as

$$\sum_{j=1}^k \sum_{\{i:\pi_i=j\}} SD(p(x|m_i, S_i), p(x|\mu_j, \sum_j)) \quad (63.8)$$

where  $p(x|m_i, S_i)$  denotes the PDF of the original  $i$ -th multivariate Gaussian and  $p(x|\mu_j, \sum_j)$  is the PDF of the final  $j$ -th multivariate Gaussian.

#### 63.3.1 Divergence Factorization

The clustering algorithm we designed depends on the elegant derivation for GMM using a class of Bregman divergence to efficiently factorize the model. Thus, the optimal parameter can be interpreted into a summation of a Mahalanobis distance and Burg matrix divergence. The two divergence measures both belong to Bregman divergence, and have closed-form optimal solutions [8].

The relative entropy between distributions can be naturally factorized into two forms:

$$D(f||g) = \int f \log f - \int f \log g = -h(f) - \int f \log g \quad (63.9)$$



The first term in (63.9) is the entropy of  $p(x|\mu, \Sigma)$  and has the following form when  $p(x|\mu, \Sigma)$  is a multivariate Gaussian.

$$h\left(p\left(x|\mu, \Sigma\right)\right) = \frac{d}{2} + \frac{1}{2} \log(2\pi)^d |\Sigma| \quad (63.10)$$

For the second term in (63.9):

$$\begin{aligned} \int p(x|m, S) \log p(x|\mu, \Sigma) &= \int p(x|m, S) \left[ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) - \log(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} \right] \\ &= -\frac{1}{2} \int p(x|m, S) \text{tr} \left( \Sigma^{-1} (x - \mu)(x - \mu)^T \right) \\ &= -\frac{1}{2} \text{tr} \left( \Sigma^{-1} E_f [(x - \mu)(x - \mu)^T] \right) - \frac{1}{2} \log(2\pi)^d |\Sigma| \\ &= -\frac{1}{2} \text{tr} \left( \Sigma^{-1} S + \Sigma^{-1} (m - \mu)(m - \mu)^T \right) - \frac{1}{2} \log(2\pi)^d |\Sigma| \\ &= -\frac{1}{2} \text{tr} \left( \Sigma^{-1} S \right) - \frac{1}{2} (m - \mu)^T \Sigma^{-1} (m - \mu) - \frac{1}{2} \log(2\pi)^d |\Sigma| \\ &\quad - \int p(x|m, S) \log(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} \end{aligned} \quad (63.11)$$

In the above derivation,  $E_f$  is the expectation for  $p(x|m, S)$ , and we use the definition  $\Sigma = E[(x - m)(x - m)^T]$  and  $E[(x - m)(m - \mu)^T] = 0$ .

Integrating (63.9) with (63.10), (63.11), we have the relative entropy

$$\begin{aligned} D(f||g) &= -\frac{d}{2} - \frac{1}{2} \log(2\pi)^d |\Sigma| + \frac{1}{2} \text{tr} \left( \Sigma^{-1} S \right) \\ &\quad + \frac{1}{2} (m - \mu)^T \Sigma^{-1} (m - \mu) + \frac{1}{2} \log(2\pi)^d |\Sigma| \\ &= \frac{1}{2} \left( \text{tr} \left( S \Sigma^{-1} \right) - \log |\Sigma| \Sigma^{-1} - d \right) + \frac{1}{2} (m - \mu)^T \Sigma^{-1} (m - \mu) \\ &= \frac{1}{2} B(S, \Sigma) + \frac{1}{2} M_{\Sigma^{-1}}(m, \mu) \end{aligned} \quad (63.12)$$

Thus the relative entropy is interpreted as a summation of Burg matrix divergence  $B(S, \Sigma)$  and Mahalanobis distance  $M_{\Sigma^{-1}}(m, \mu)$ . For symmetric relative entropy measure, we have

$$SD(f, g) = B(S, \Sigma) + \frac{1}{2} \{ M_{\Sigma^{-1}}(m, \mu) + M_{S^{-1}}(\mu, m) \} \quad (63.13)$$

The problem for clustering is to find an optimal parameter  $\mu^*$ ,  $\Sigma^*$  for  $k$  ensemble clusters.

$$\begin{aligned}
 p(x|\mu^*, \sum^*) &= \arg \min_{p(x|\mu, \sum)} \sum_i SD(p(x_i|m_i, S_i), p(x|\mu, \sum)) \\
 &= \arg \min_{p(x|\mu, \sum)} \sum_i B(S, \sum) + \frac{1}{2} \{M_{\sum^{-1}}(m_i, \mu) + M_{S^{-1}}(\mu, m_i)\}
 \end{aligned}
 \tag{63.14}$$

The optimal  $\sum^*$  can be derived by setting the gradient of (63.14) to 0 with respect to  $\sum^{-1}$ .

$$\begin{aligned}
 \frac{\partial}{\partial \sum^{-1}} \sum_i SD(p(x_i|m_i, S_i), p(x|\mu, \sum)) &= \sum_{i=1}^n (S_i - \sum + (m_i - \mu^*)(m_i - \mu^*)^T) \\
 &= 0
 \end{aligned}
 \tag{63.15}$$

Consequently,

$$\sum^* = \sum_i (S_i + (m_i - \mu^*)(m_i - \mu^*)^T)
 \tag{63.16}$$

For the optimal  $\mu^*$ , a unique minimizer is provided in [6] with

$$\mu^* = \sum_i m_i
 \tag{63.17}$$

### 63.3.2 Algorithm

We present our clustering algorithm for GMM below. The algorithm works in iterative manner and the optimal parameter for final ensemble clusters are shown in (63.16), (63.17).

Information-theoretical Clustering Algorithm

**Step 1.** Initialize original n multivariate Gaussians

$$\{m_1, \dots, m_n\}, \{S_1, \dots, S_n\}$$

**Step 2.** Assign the n Gaussians into k Clusters







$$\pi_i \in \{1, 2, \dots, k\}, i = 1, \dots, n$$

**Step 3.** Iterative optimize the parameter

**while** not converged do

**for** j = 1 to k

**Table 63.1** Application of information-theoretic clustering algorithm to clustering-based image segmentation

Algorithm	Num. of clusters = 3	Num. of clusters = 5
k-means		
GMM-EM		
Information-theoretic Clustering		

$$\sum_j^* = \sum_i (S_i + (m_i - \mu^*)(m_i - \mu^*)^T)$$

**end for**  
**for** i = 1 to n

$$\pi_i = \arg \min_{1 \leq j \leq k} B(S_i, \sum_j) + M_{\sum_j^{-1}}(m_i, \mu_j)$$

**end for**  
**end while**

Note that, for sensor network data and certain application where data is spatially diffused, modeling of the nodes and initialization of the algorithm is simple. While for other application such as clustering-based image segmentation, the estimation of original multivariate Gaussians is critical to the final result.

**Table 63.2** Runtime (in seconds) comparison among different clustering algorithms

	3 clusters	5 clusters	10 clusters
k-means	1.23	1.51	2.48
GMM-EM	2.33	3.17	5.32
<b>Information-theoretic Clustering</b>	<b>1.81</b>	<b>2.79</b>	<b>4.74</b>

## 63.4 Experiments

Our algorithm is demonstrated in practical problem, clustering-based image segmentation [3, 9], with comparison to conventional k-means and GMM-EM. The effectiveness of our algorithm is superior to k-means and GMM-EM directly by visual effects. Also, the computational cost of our algorithm is satisfactory evaluated by the runtime.

Given an RGB image, a pixel  $x$  can be considered as a point in  $R^3$ . We can model the segmentation using GMM to describe the separate components. In experiment, we extract 11 features for the input color image, including RGB, Luv, Lab, x-axis, y-axis, gray etc.

Table 63.1 demonstrates the superiority of our information-theoretic clustering algorithm. In comparison to k-means and GMM-EM, the algorithm performs the best visual effect in column, and its robustness is shown on dealing with different demands on number of clusters in row.

Table 63.2 compares our algorithm with two conventional cluster algorithms. Synthesis dataset used in [3] is employed to provide a standard evaluation for the algorithms. The codes are written in MATLAB and run on a 2.93 GHz Core-Duo machine.

## 63.5 Conclusions

We use symmetric relative entropy and factorize the multivariate GMM aiming at efficiently reducing the number initialized clusters. The divergence of distributions can be interpreted into a summation of two components of Bregman divergence and the optimal parameter can be then solved by simple optimization manipulation. The algorithm demonstrates better visual effects in comparison to k-means and GMM-EM at relatively lower computational cost as a simplification of GMM.

**Acknowledgments** This work is supported by National Natural Science Foundation (NNSF) of China under Grant No. 61004023, No. 60004021, No. 61074057.

## References

1. Duda RO, Hart PE, Stork DG (2001) Pattern classification. Wiley, New Jersey
2. Ng A, Jordan M, Weiss Y (2002) On spectral clustering: analysis and an algorithm. In: Proceedings of the 16th annual conference on neural information processing systems, pp 849–856
3. Zhang K, Kwok JT (2006) Simplifying mixture models through function approximation. In: Proceedings of the 20th annual conference on neural information processing systems, pp 1577–1584
4. Goldberger J, Greenspan H, Dreyfuss J (2007) Simplifying mixture models using the unscented transform. *IEEE Trans Pattern Anal Mach Learn* 30(8):1496–1502
5. Nielsen F, Garcia V, Nock R (2009) Simplifying Gaussian mixture models via entropic quantization. In: Proceedings of the 17th European signal processing conference (EUSIPCO 2009), pp 2012–2016
6. Banerjee S, Merugu ID, Ghosh S (2005) Clustering with Bregman divergences. *J Mach Learn* 6:1705–1749
7. Goldberger J, Gordon S, Greenspan H (2003) An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures. In: Proceedings of the IEEE international conference on computer vision, pp 172–177
8. Davis JV, Dhillon I (2006) Differential entropic clustering of multivariate Gaussians. In: Proceedings of the 20th annual conference on neural information processing systems, pp 552–559
9. Borenstein E, Ullman S (2004) Learning to segment. In: Proceedings of the 8th European conference on computer vision (ECCV), pp 235–232

# Chapter 64

## A Method of Acceleration Applied in Symmetric-SIFT and SIFT

Dong Zhao, Qi Wang, Haiyan Sun and Xiaopeng Hu

**Abstract** Symmetric-SIFT is an effective technique used for registering multimodal images. It is based on a well-known image registration technique named Scale Invariant Feature Transform (SIFT). Similar to SIFT, Symmetric-SIFT detects many stable keypoints even though not all of which are useful. Experiments show that matching keypoints are mostly on or near the edge. Based on the phenomenon, we propose an effective method. In our method, We extract the edge and classify keypoints by whether they are on the edge or not. Then we delete the points that are far from the edge and match the remained ones. Finally, we get the matching set after filtering the initial matching result with the threshold value got by the Bayesian formula. The experimental results show that the proposed method can not only greatly reduce the matching time, but also effectively improve the matching rate.

**Keywords** Multimodal image registration · Symmetric-SIFT · Edge · Bayesian

### 64.1 Introduction

Image registration is a fundamental aspect of many problems in computer vision, including object or scene recognition, solving for 3D structure from multiple images, stereo correspondence, and motion tracking. The purpose of image registration is to find the appropriate geometric transformation required to map the sensed image on the reference image. In 2004, David Lowe [1] proposed a novel point characteristics extracting method called SIFT. The SIFT algorithm can

---

D. Zhao (✉) · Q. Wang · X. Hu  
Faculty of Electronic Information and Electrical Engineering,  
Dalian University of Technology, Lingshui Road, Dalian, China  
e-mail: zddlutstone@yahoo.com.cn

H. Sun  
HONGDU Aviation Industry Group LTD., AVIC. JianXi, NanChang, China

generate a large number of distinguished keypoints. However, it consumes plenty of time and doesn't meet real-time conditions. In order to accelerate the implementation, many researchers have proposed lots of improved algorithms of SIFT, such as PCA-SIFT [2], GLOH [3], SURF [4], etc.

Two or more images are called multimodal if each of them is captured by a different sensor (imaging device or modality) [5]. Different sensors with different levels of sensitivity to a particular part of an object results in the same portion of an object represented by different intensities in images captured from different modalities [6]. Variation in intensities has a possible consequence 'Gradient Reversal' [5], namely in multimodal images the gradients of corresponding parts of the images will change their directions by exactly  $180^\circ$  [6, 7]. Gradient Reversal invalidate SIFT for multimodal images registration. The paper [8] puts forward a novel technique named Symmetric-SIFT which has been adapted to multimodal registration. Similar to SIFT, Symmetric-SIFT generates a large number of keypoints, but simultaneously consume plenty of time.

In this paper, we propose a method to speed up SIFT and Symmetric-SIFT, which can effectively improve the matching rate and greatly shorten the matching time. Our experiments' purpose is processing multimodal images, so in our paper the proposed method is mainly told based on Symmetric-SIFT, which doesn't affect the method's applicability to SIFT. The experiments show that the proposed method can achieve the expected results.

The rest of the paper is organized as follows. Section 64.2 briefly introduces the original SIFT and Symmetric-SIFT. Section 64.3 describes our proposed method. Section 64.4 presents experimental results. Finally Sect. 64.5 concludes this paper.

## 64.2 Overview of SIFT and Symmetric-SIFT

### 64.2.1 SIFT

The SIFT algorithm introduces the concept of multi-scale space [9–11], which can make the keypoints with scale information. It employs difference-of-Gaussian (DOG) function to approximate Laplacian of Gaussian operator. Comparing adjacent three layer DOG images we can get the extreme value points. Then least square fitting is conducted by Taylor expansion of scale-space function to eliminate the unstable extreme-points.

After extracting keypoints, a main orientation need be assigned to each keypoint in order to achieve invariance to image rotation. An orientation histogram is formed from the gradient orientations of sample points within a region around the keypoint. Peaks in the orientation histogram are considered as dominant directions of keypoints. Any other local peak that is within 80 % of the highest peak is also used to create a keypoint with the orientation.

Next step is to compute a descriptor of each keypoint based on the local image region. The local region is divided into  $16 \times 16$  sample regions. After weighted

with a Gaussian window, eight gradient direction histograms could be obtained in each  $4 \times 4$  subregion. So this will generate a 128 dimensional vector. Finally, the feature vector need be normalized to make it invariant to illumination changes.

After the feature vectors of two images are generated, the Euclidean distance of the keypoints feature vectors is used to measure the similarity of keypoints between two images.

### 64.2.2 Symmetric-SIFT Algorithm

Compared to SIFT, Symmetric-SIFT [8] has two steps that are different: orientation assignment and keypoint descriptor.

#### (a) Symmetric Orientation Assignment

Symmetric-SIFT settles the gradient reversal problem by limiting all the gradients within the range  $[0, \pi)$ . Symmetric-SIFT applies a continuous method to assign the main orientation to each keypoint instead of the discrete one of SIFT based on the gradient histograms. Then to avoid opposite gradient vectors canceling each other when gradients are averaged directly, squaring the gradient vector is proposed before averaging. The operation of the above is for each sample point in a region around the keypoint location, and next the average squared gradient vector can be calculated by averaging the neighborhood convolved by the Gaussian-weighted circular window. Finally, the dominant direction  $\varphi$  of each neighborhood can be determined using the smoothed vector.

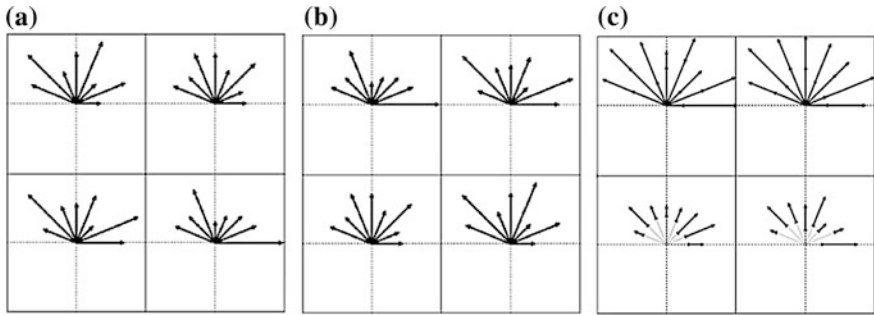
#### (b) Symmetric Keypoint Descriptor

The directions of gradients in this step are also restricted from 0 to  $\pi$ . Eight gradient direction histograms in each  $4 \times 4$  subregion could be obtained as shown on Fig. 64.1a. Rotating 180 degree of the original neighborhood around the keypoint, we can then obtain another eight gradient direction histograms as shown on Fig. 64.1b. In this case, we can get two 128-dimension feature vectors. Finally, the elements of top half descriptor are the sum of the corresponding elements of the two 128-dimension feature vectors, and the elements of bottom half descriptor are the absolute values of difference of the corresponding elements of the two vectors as shown on the right of Fig. 64.1. Now an effective descriptor for multimodal images registration can be obtained.

## 64.3 Proposed Technique to Accelerate Registration Speed

As previously mentioned, redundant keypoints reduce the matching rate, and greatly consume the matching time. Through experiments we obtain that most of the successful matching points distributed on or near the edge. Using this





**Fig. 64.1** Symmetric-SIFT descriptor

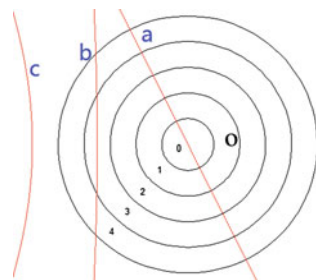
information, we mark and screen the keypoints, to speed up the matching time and improve the matching rate. After that, we screen the matching points through Bayesian formula to improve the number of correct matching.

First, we get an edge image through the edge detector. Here we choose Sobel operator to extract the edge, in which we set the threshold  $T$  as 0.1 (unitized).

Then we need measure the minimum distance  $d$  between the extracted keypoints and the edge. But the edge can't be represented by a function, which leads to the fact that the distance can't be measured directly. Moreover, the edge extracted by the edge detection algorithm is not accurate. Therefore we use a method of intersection with concentric circles to imitate the processing. The method is as follows: a series of concentric circles can be drawn using the keypoints as the center and 2, 4, 6, 8, 10 as the radius, and the minimum number of the concentric circles which intersect with the edge is regarded as the minimum distance  $d$ , as shown in Fig. 64.2.

In this way, we will classify all keypoints into three categories. The first category is the keypoints on the edge including edge nearest neighbors, and in this case we set the  $d$  as 0. The second category is the keypoints far from the edge, namely the keypoints outside the circle which uses the keypoint as the center and 10 as the radius. We delete the keypoints included in the second category for the little contribution they offer. The third category is the keypoints near the edge which are between the first and the second category. We can obtain the distance  $d$  through the previously mentioned method.

**Fig. 64.2** Line a, b, c refer to respectively three edges, the center O indicates the keypoint and circles with radius 2, 4, 6, 8, 10 correspond respectively to distance numbers 0, 1, 2, 3, 4



In the matching stage, the edge points ( $d = 0$ ) are matched to the edge points, the non-edge points ( $d \neq 0$ ) are matched to the non-edge points. According to the different values of  $d$ , we dynamically set the threshold of the ratio between the distance of the closest neighbor and that of the second-closest neighbor. The greater is the  $d$  value, the lower is the threshold value set. As a result, we will draw the initial matching results.

We define one keypoint's  $d$  as its distance from the edge, and one match's  $d_{\text{difference}}$  as the difference of the two matching points'  $d$ s. In those matches, the bigger the  $d_{\text{difference}}$  of one match is, the smaller the possibility of the match's successful matching is. In order to guarantee high matching rate, aiming at the matching results of non-edge points, we use the threshold value estimated by Bayesian formula to screen the results. First of all through the statistics of the matching results of non-edge points, the probabilities  $f\langle d_{\text{difference } i} | S \rangle$  and  $f\langle d_{\text{difference } i} | F \rangle$  of different  $d_{\text{difference}}$  in the correct matching and wrong matching can be obtained, in which  $d_{\text{difference}}$  is 0,1,2,3. And we also can get the probabilities  $f(S)$  and  $f(F)$  of the correct and wrong matches. Then the correct matches probabilities of different  $d_{\text{difference}}$  can be calculated according to (64.1). Through experiments we get  $f\langle S | d_{\text{difference}} = 0 \rangle = 0.98$ ,  $f\langle S | d_{\text{difference}} = 1 \rangle = 0.90$ ,  $f\langle S | d_{\text{difference}} = 2 \rangle = 0.85$ ,  $f\langle S | d_{\text{difference}} = 3 \rangle = 0.80$ . By the analysis of above data, this article chooses  $d_{\text{difference}} = 1$  as the threshold value to filter the results, namely that if  $d_{\text{difference}} < 1$ , we retain the matches, otherwise delete it. In this way, we will obtain the better matching results.

$$f\langle S | d_{\text{difference}} \rangle = \frac{f\langle d_{\text{difference}} | S \rangle f(S)}{f\langle d_{\text{difference}} | S \rangle f(S) + f\langle d_{\text{difference}} | F \rangle f(F)} \quad (64.1)$$

## 64.4 Experiments and Results

The development environment of the experiment is Intel Core 2 2.94 GHz CPU, 2G Memory. The operating system is 64 bit Window 7. The development platform is Visual Studio 2010 and OpenCV2.3. The database images which are used in experiments contain 30 pairs of multimodal images and color images from different sources [1, 3, 8, 12].

To evaluate the proposed method, we need define two evaluation measure: matching rate  $P$  and acceleration ratio  $T$ . The specific calculations are according to (64.2) and (64.3) respectively.

$$P = \frac{2 * \text{Correct Matches}}{\text{Keypoints1} + \text{Keypoints2}} * 100\% \quad (64.2)$$

$$T = \frac{t_{\text{original}} - t_{\text{proposed}}}{t_{\text{original}}} * 100\% \quad (64.3)$$

We verify the effectiveness of the proposed method from the matching rate  $P$  and acceleration ratio  $T$ .

*(a) The Matching Rate  $P$*

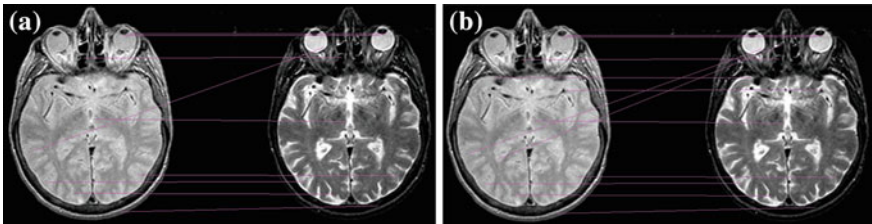
As shown in Fig. 64.3, the Symmetric-SIFT algorithm totally generates 1923 keypoints and the matching rate  $P$  calculated by (64.2) is 0.83 %. While the proposed method generates 1900 keypoints and the matching rate  $P$  calculated by (64.2) is 1.05 %, which is higher than the Symmetric-SIFT algorithm. We also use another 4 pairs of multimodal images to test. Table 64.1 clearly shows the new method effectively improves the matching rate  $P$ .

*(b) The Acceleration Ratio  $T$*

Due to separate matching between edge points and non-edge points, the registration time is greatly shortened. As shown in Table 64.2, the maximum of acceleration ratio reach up to 85.5 %, which implies the registration time is reduced by 85.5 %.

Then we also apply the proposed method to the SIFT algorithm, as shown in Fig. 64.4. From the contrast figures we can find the matching rate of keypoints detected by the proposed method increases by 3.07 %, and the acceleration rate reaches up to 67.3 %.

Now let us examine why the proposed method can achieve desired effects. The points far from the edge contribute little to the matching result, so deleting them can enhance the matching rate. Then edge points and non-edge points separately matching will increase the matching reliability. And more importantly it greatly reduces the matching time, because every time matching keypoints don't need to traverse all keypoints. At last the initial matching result will be screened to increase the proportion of the correct match. Thus we can achieve the purpose of design in this paper.



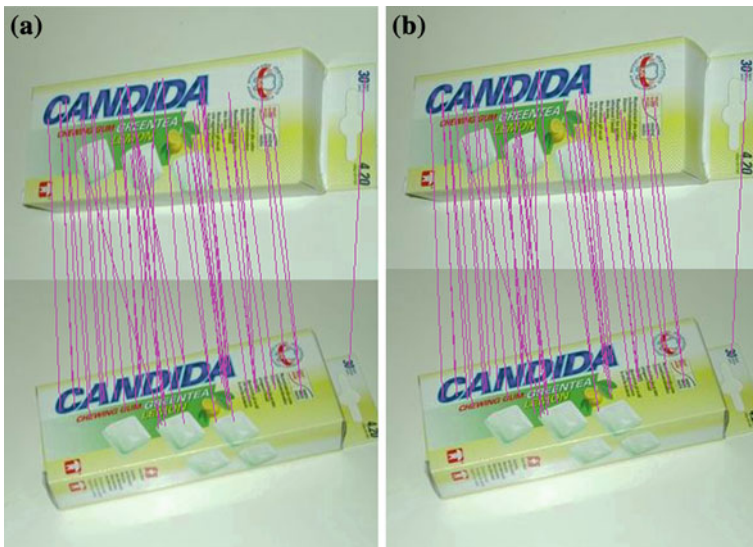
**Fig. 64.3** Registration results with Symmetric-SIFT and the proposed method. **a** Symmetric-SIFT: 1923 keypoints in total, 8 pairs of correct matches and consuming 1035.6 ms. **b** the proposed method: 1900 keypoints in total, 10 pairs of correct matches and consuming 620.3 ms

**Table 64.1** Comparison of our proposed method and Symmetric-SIFT for the matching rate P

Specimen	Type of image pair	Matching Rate P	
		Symmetric-SIFT (%)	The proposed metho (%)
Brain3	MR T1-T2	0.85	1.15
city	EO/IR	0.9	1.51
Brain1	MR T1-T2	0.14	0.69
EO-IR-2	EO/IR	0.58	0.84

**Table 64.2** Comparison of our proposed method and Symmetric-SIFT for the acceleration ratio T

Specimen	Type of image pair	RUNTIME (ms)		
		Symmetric-SIFT	The proposed method	The acceleration ratio T (%)
Brain3	MR T1-T2	1121.1	718.5	35.9
City	EO/IR	648.4	450.2	30.6
Brain1	MR T1-T2	553	275.2	50.2
EO-IR-2	EO/IR	9997.1	1454.4	85.5



**Fig. 64.4** Registration results with SIFT and the proposed method. **a** SIFT: 409 keypoints in total, 41 pairs of correct matching points and consuming 233.4 ms. **b** Improved-SIFT: 398 keypoints in total, 46 pairs of correct matching points and consuming 76.4 ms

## 64.5 Conclusion

In the paper, we propose using the edge information to classify the keypoints, to enhance the matching rate and reduce the runtime. The same to SIFT, Symmetric-SIFT can generate many useless keypoints. We delete the keypoints far from edges, then respectively match the rest keypoints of different categories and finally filter the matching results. Experiments results show that the proposed method is effective for Symmetric-SIFT and SIFT, especially in accelerating registration speed.

**Acknowledgments** This research was supported by “National Natural Science Foundation of China” (No.61272523) and “the National Key Project of Science and Technology of China” (No.2011ZX05039-003-4).

## References

1. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis.* 60(2):91–110
2. Ke Y, Sukthankar R (2004) PCA-SIFT: a more distinctive representation for local image descriptors. In: *Proceedings of IEEE conference on computer vision and pattern recognition*, pp 506–513
3. Mikolajczyk K, Schmid C (2005) A performance evaluation of local descriptors. *IEEE Trans Pattern Anal Mach Intell* 27(10):1615–1630
4. Bay H, Tuytelaars T, Gool LV (2006) SURF: speeded up robust features. In: *European conference on computer vision*, pp 404–417
5. Hossain T, Lv G, Teng SW, Lu G, Lackmann M (2011) Improved Symmetric-SIFT for multi-modal image registration. In: *2011 international conference on digital image computing: techniques and applications*, pp 197–202
6. Pluim J, Maintz J, Viergever M (2000) Image registration by maximization of combined mutual information and gradient information. In: *medical image computing and computer assisted intervention-MICCAI 2000*. Springer, Berlin, pp 103–129
7. Collignon A, Vandermeulen D, Suetens P, Marchal G (1995) 3D multi-modality medical image registration using feature space clustering. In: *Computer Vision, Virtual Reality and Robotics in Medicine*. Springer, Berlin, pp 193–204
8. Chen J, Tian J (2009) Real-time multi-modal rigid registration based on a novel symmetric-SIFT descriptor. *Prog Nat Sci* 19(5):643–651
9. Koenderink JJ (1984) The structure of images. *Biol Cybern* 50:363–396
10. Lindeberg T (1994) Scale-space theory: a basic tool for analyzing structures at different scales. *J Appl Stat* 21(2):224–270
11. Witkin, A.P. (1983). Scale-space filtering. In: *International joint conference on artificial intelligence*, Karlsruhe, Germany, 1019–1022
12. Kelman A, Sofka M, Stewart C (2007) Keypoint descriptors for matching across multiple image modalities and non-linear intensity variations. In: *2007 IEEE conference on computer vision and pattern recognition*, pp 1–7

# Chapter 65

## A Novel Method of Image Enhancement via Multi-Scale Fuzzy Membership

Ce Li, Yannan Zhou and Chengsu Ouyang

**Abstract** For the imaging condition restriction, nature images sometimes have the problem with low contrast and low illumination. In order to solve those problems, we proposed a novel image enhancement algorithm based on multi-scale fuzzy membership. The images will be firstly decomposed into multi-scale sub images by Laplacian pyramid and then be enhanced through calculating the fuzzy membership degree under multi-scale. Finally the edge-preserving image denoising will be conducted by bilateral filter to realize the effectively enhancement of the nature images. The experimental result shows that the proposed algorithm also can be used in X-ray medical image enhancement and achieves a better effect compared with the traditional method and has certain theoretical and practical application value.

**Keywords:** Image enhancement · X-ray Image · Multi-scale · Fuzzy membership · Bilateral filter

### 65.1 Introduction

Because the restriction of imaging condition such as backlighting, evening or night photography and so on, nature images always with low quality, similarly in the X-ray medical imaging system. Due to the characteristics like low ray's dosage and biological tissue thickness not equal [1, 2]. The X-ray medical images often appear low contrast and low illumination needs to be enhanced, too. But the traditional image enhancement method shows a lot of limitations. For example, classic

---

C. Li (✉) · Y. Zhou  
College of Electrical and Information Engineering, Lanzhou University of Technology,  
Lanzhou 730050, China  
e-mail: xjtulice@gmail.com

C. Li · C. Ouyang  
The School of Electronic and Information Engineering, Xi'an Jiaotong University,  
Xi'an 710049, China

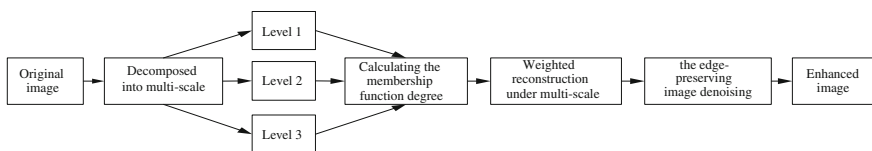
algorithms such as Prewitt and Soble [3] often use gradient factor for image enhancement. Prewitt also enhanced image noise while enhancing the useful information, and Soble got disadvantage like rough edges and sensitive to noise. The gray level may be over integrated when conducting image enhancement through histogram equalization method and contrast gain method [4]. The traditional image enhancement method just simply change the contrast of the image or suppress the noise without taking the fuzziness into account [5], which usually weaken the details of the image when suppress the noise. So Pal and King [6] proposed a fuzzy image enhancement algorithm. The algorithm keeps the detail information of the image when enhancing the original images. Meanwhile, this algorithm did not take the multi-scale and the edge-preserving noise denoising of the visual target.

In order to solve the above problems, in this paper, a novel method for image enhancement is proposed, which is based on multi-scale fuzzy membership. Our experimental results show that the proposed method achieves better performance in image enhancement, especially when the images are extremely low contrasted and low illuminated.

This paper is organized as follows: Sect. 65.2 describes the proposed method and concrete steps, including the Laplacian pyramid, calculating the fuzzy membership degree, weighted reconstruction under multi-scale and the edge-preserving image denoising by bilateral filter. Section 65.3 presents the experimental results obtained using the proposed method and other existing method. Finally, conclusions will be drawn in Sect. 65.4.

## 65.2 Proposed Algorithm

The frame of our proposed algorithm is shown in Fig. 65.1. In our proposed method, firstly, the original image will be decomposed into multi-scale sub images by the Laplacian pyramid. According to the characteristics of the multi-scale visual perception of human eyes we know that sub images under different scale contain different detail information. Then the decomposed images at each level will be enhanced in detail through calculating the fuzzy membership degree. According to the calculated the fuzzy membership degree of each level, add different weights to each sub image and reconstruct. Finally obtain the enhanced image. At the end, for images usually have the default of edge blur, the edge-preserving image denoising will be conducted by bilateral filter.



**Fig. 65.1** The frame of our proposed algorithm

### 65.2.1 The Laplacian Pyramid

In Human visual simulation, multi-scale pyramid is the main form of demonstrating the image in multi-scale. Image pyramid algorithm was proposed by Burt and Anderson [7]. The raw image can be decomposed into different scales of sub image by this algorithm, each sub image include detail information in their scale. The composition of the Laplacian pyramid is evolved on the basis of the Gaussian pyramid, each level of the Laplacian pyramid represents the difference between successive levels of the Gaussian pyramid, reflects the information gap between two levels of the Gaussian pyramid, also is the detail part of the image.

### 65.2.2 Fuzzy Membership

The image from the structure can be divided into regions of interest and background. Generally speaking, the pixels with high gray value belong to information regions, pixels with low gray value belong to background regions. We do not have a fixed gray value as the criterion between these two regions. In this paper, an image  $I$  of size  $M \times N$  pixels, having  $L$  gray-levels ranging from  $L_{\min}$  to  $L_{\max}$ , can be viewed as an array of fuzzy singletons [8]. Image  $I$  can be described as:

$$I = \{\lambda_I(g_{ij})/g_{ij}, i = 1, 2, \dots, M, j = 1, 2, \dots, N\} \quad (65.1)$$

In image processing, we use the bi-level threshold method used to classify pixels to dark group or bright group [9]. With this goal, two fuzzy sets (dark and bright) may be considered. Whose membership functions  $\lambda_{dark}$  and  $\lambda_{bright}$  are defined as below:

$$\lambda_{dark} = \begin{cases} \frac{1}{\alpha}y^2, & \text{if } y \in [0, \alpha], \\ 0, & \text{if } y \in [\alpha, 1], \end{cases} \quad (65.2)$$

$$\lambda_{bright} = \begin{cases} 0, & \text{if } y \in [0, \alpha], \\ \frac{1 - B(y)}{1 + \beta B(y)}, & \text{if } y \in [\alpha, 1], \end{cases} \quad (65.3)$$

where

$$B(y) = \frac{1}{\alpha} \left( \frac{1 - y}{1 + \beta y} \right)^2 \quad (65.4)$$



In this,  $y \in I$  is the independent variable, and  $\alpha \in (0, 1)$  is the crossover point. The relationship of  $\alpha$  and  $\beta$  is as below:

$$\beta = \frac{1 - 2\alpha}{\alpha^2} \tag{65.5}$$

The parameter  $\varepsilon$  is a positive threshold value of fuzzy sure entropy. In our approach, the experimental results proved that the value of  $\varepsilon$  directly connect with the image enhancement performance. Here, for selecting a suitable  $\varepsilon$ , we firstly define the first-order fuzzy moments  $m$  and  $P(I_c)\max$  ( $c$  is the number of partitions) as follows:

$$m = \sum_{g=0}^{L-1} t(g) \tilde{h}_I(g) \tag{65.6}$$

and

$$t(g) = \frac{g}{L - 1} \tag{65.7}$$

$$\tilde{h}_I(g) = \frac{h_I(g)}{M \times N} \tag{65.8}$$

where  $L$  is gray levels,  $\tilde{h}_I(g)$  denotes the probability of the normalized histogram  $h_I(g)$  corresponding to gray-level  $g$  of the image  $I$ . In our study,  $P(dark)_{\max}$  and  $P(bright)_{\max}$  are defined as the maximal values of  $P(dark)$  and  $P(bright)$ . Then they are:

$$P(dark)_{\max} = \max(P_\alpha(dark)|\alpha \in (0, 1)) \tag{65.9}$$

and

$$P(bright)_{\max} = \max(P_\alpha(bright)|\alpha \in (0, 1)) \tag{65.10}$$

Here, we can assign the value of  $\varepsilon$  as below:

$$\varepsilon = \begin{cases} \frac{P(bright)_{\max}}{2}, & \text{if } m \leq 0.5, \\ \frac{P(dark)_{\max}}{2}, & \text{if } m > 0.5, \end{cases} \tag{65.11}$$

When  $m$  is less than 0.5, the image looks dark, then we can select  $\varepsilon$  ranged from 0 to  $P(bright)_{\max}$  in order to obtain a satisfying image. Here, the greater  $\varepsilon$  is, the brighter the enhanced image looks. However, when  $\varepsilon$  is more than  $P(bright)_{\max}$ , the enhanced image no longer alters along with  $\varepsilon$ 's change. On the other hand, when  $m$  is more than 0.5 and the image looks bright, we select  $\varepsilon$  ranged from 0 to  $P(bright)_{\max}$ . Moreover, the less  $\varepsilon$  is, the darker the enhanced image looks.

### 65.2.3 Weighted Reconstruction Under Multi-Sale

Each sub image decomposed by the Laplacian pyramid represents different detail information. In this paper, images will be enhanced through calculating the fuzzy membership degree of each sub image and weighted reconstruction under multi-scale. After the enhancement under multi-scale, each level sub image was weighted by follows:

$$\bar{I} = \frac{1}{L} \sum_{l=0}^{L-1} \sum \alpha_l \varepsilon_l I_l, \quad l = 0, 1, \dots, L-1 \quad (65.12)$$

where,  $L$  is the number of decomposed levels. In our proposed method,  $L = 3$ . Here, weighted coefficient  $\alpha_1$  should be  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.3$ ,  $\alpha_3 = 0.6$ . Finally, we reconstructed the image after weighted.

### 65.2.4 Bilateral Filter

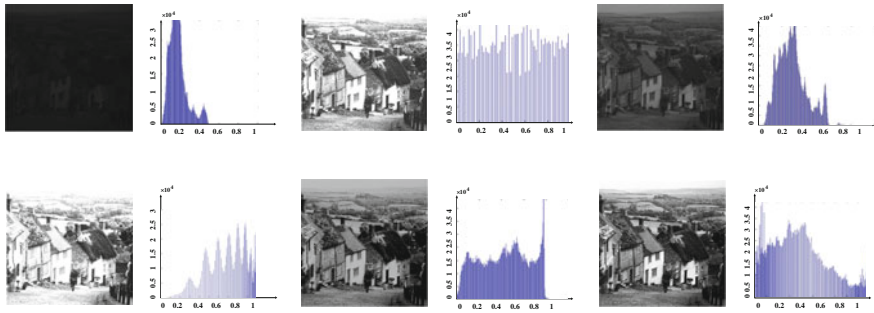
The concept of bilateral filter is first put forward by Tomasi and Manduchi [10] in 1998. It is a nonlinear filtering algorithm and its fundamental is proposed based on Gaussian filter. When dealing with gray value in the pixel neighborhood, bilateral filter does not only consider the distance relationship of the position in space, but also take into consideration the distance relationship between adjacent pixel gray values. Then denoise the image through nonlinear combination of the two. The bilateral filter consists of a domain filter and a spatial domain filter. The definition of bilateral filter is as follows:

$$h(x) = k^{-1}(x) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\xi) c(\xi, x) s(f(\xi), f(x)) d(\xi) \quad (65.13)$$

From the definition of bilateral filter we can find that it enhances not only the smoothness of the geometrical domain, but also that of the brightness domain. In smooth regions, pixel values in a small neighborhood are similar to each other. In that case, the effect of bilateral filter enhancement is nearly the same as pure spatial domain filter when coming into the edge of the image. The feature of photometric similarity equation ensures the effect of the edge-preserving image denoising during the process of image enhancement through bilateral filter.

## 65.3 Experimental Results

In this section, in order to evaluate the performance of the proposed method, we first tested the proposed method using nature images with low contrast and low luminance. Then we tested X-ray images from DDR system, and compared our



**Fig. 65.2** The original image with low contrast and low luminance enhanced by our proposed method and other four methods. The first line from *left to right* is the original nature image with low luminance and low contrast, the enhanced image obtained by HEM and MF. The second line from *left to right* is the enhanced image obtained by PIFM [5], MSNM [12] and our proposed method. Each method’s histogram is obtained as well

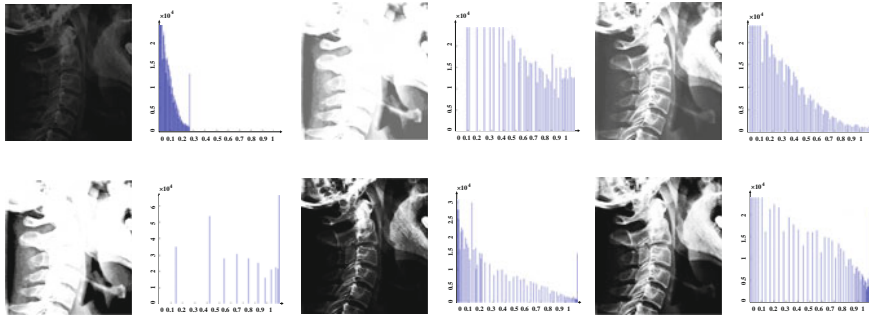
results with those produced method using the histogram equalization method (HEM), the median filter ( $3 \times 3$ ) (MF), the maximizes parametric index of fuzziness (PIFM) and the maximum Shannon entropy principle of fuzzy events (MSNM). The experimental environment is MATLAB 7.0, CPU is Intel Core i5 2.27 GHz, RAM is 2G (Fig. 65.2).

Then we compared the performance of image quality evaluation on our proposed and other four methods, the results is shown in Table 65.1. In this paper, we not only test the man square error (MSE) and peak signal to noise ratio (PSNR) of traditional image quality evaluation, but also we test the structural similarity (SSIM) [11] to be consistent with human eye perception. From the result of Table 65.1, it shows that the proposed algorithm of image enhancement is better effect compared with the traditional method.

From Fig. 65.3, we can see the image obtained by HEM and PIFM [5] all have some over-enhanced portions which make them unclear and unnatural. And the image obtained by MF and MSNM are under-enhanced. In this test, the proposed method gets ahead of the other methods. The image enhanced by the proposed method looks not only distinct but also. Similarly, we can see the histogram of results, that our proposed method is better compared to other algorithms.

**Table 65.1** Performance comparison of the enhanced image quality evaluation on nature image obtained by our proposed and other four methods

	HEM	MF	PIFM	MSNM	Proposed
MSE	1.4458e + 003	1.9994e + 003	7.2265e + 003	4.0154e + 03	0.7004e + 03
PSNR	16.5298	15.1218	9.0600	17.5463	19.9826
SSIM [11]	0.6238	0.6919	0.4758	0.7098	0.7879



**Fig. 65.3** The comparison of our proposed method and other four methods. The first line from *left to right* is the original X-ray image ‘neck’ with low luminance and low contrast, the enhanced image obtained by HEM and MF. The second line from *left to right* is the enhanced image obtained by PIFM [5], MSNM [12] and our proposed method. Each method’s histogram is obtained as well

## 65.4 Conclusions

This paper proposes a novel developed image enhancement method based on multi-scale fuzzy membership. So the contribution of this paper is twofold. First, it is very efficient and effective when applying our proposed method on low-quality images, especially on low-luminance and low-contrast X-ray images. Second, the proposed method is quite robust as the threshold value can be selected in a relatively very large range until satisfying results. The comparisons of those experimental results show that the proposed method is better performance over the traditional algorithms.

**Acknowledgments** This work was supported by the Natural Science Foundation in Gansu province Grant Nos. (1112RJZA033, 2011GS04165). We would like to thank the anonymous reviewers for their comments.

## References

1. Panetta K, Wharton E, Agaian S (2008) Human visual system-based image enhancement and logarithmic contrast measure. *IEEE Trans Syst Man Cybern B* 38(1):174–188
2. Cheng HD, Xu H (2000) A novel fuzzy logic approach to contrast enhancement. *Pattern Recogn* 33:799–819
3. Gonzalez RC, Woods RE (2006) *Digital image processing*, 3rd edn. Prentice-Hall, Upper Saddle River
4. Xiaolin Wu (2011) A linear programming approach for optimal contrast-tone mapping. *IEEE Trans Image Process* 20(5):1262–1272
5. Ioannis KV, George DS (2006) Parametric indices of fuzziness for automated image enhancement. *Fuzzy Sets Syst* 157:1126–1138

6. Pal SK, King RA (1983) On edge detection of X-Ray images using fuzzy sets. *IEEE Trans PAMI* 25(1):69–77
7. Burt PJ, Anderson EH (1983) The Laplacian pyramid as a compact image code. *IEEE Trans Commun* 2(7):532–540
8. Tizhoosh HR, Krell G, Michaelis B (1998)  $\lambda$ -Enhancement: contrast adaptation based on optimization of image fuzziness. In: *Proceeding of the FUZZ-IEEE98*, pp 1548–1553
9. Avci E, Avci D (2009) An expert system based on fuzzy entropy for automatic threshold selection in image processing. *Expert Syst Appl* 36(2):3077–3085
10. Tomasi C, Manduchi R (1998) Bilateral filtering for gray and color images. In: *Proceeding of international conference on computer vision*, pp 839–846
11. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error measurement to structural similarity. *IEEE Trans Image Process* 13(1):600–612
12. Cheng HD, Chen JR (1997) Automatically determine the membership function based on the maximum entropy principle. *Inf Sci* 96:163–182

## Chapter 66

# Adaptive Region Clustering in LDA Framework for Image Segmentation

Xiaoru Wang, Junping Du, Shuzhe Wu and Fu Li

**Abstract** Image segmentation based on low-level features has been studied for many years. However, because of the semantic gap issue, it is difficult to have more breakthroughs based on low-level features. LDA is a powerful tool to model co-occurrence relationships between words and thus is used to catch the semantic connections between low-level visual features. In image segmentation, the codebook is built from the visual features and topics are trained with LDA model. And the topic distributions yield important guidance for segmentation. However, in previous papers, researchers used the topic with the highest probability to merge the regions. It ignored the statistics nature of the topic distribution. And, the segmentation result will be greatly impact by the codebook size, the topic number and cluster number. To address these challenges, this paper proposes a new image segmentation algorithm based on LDA framework: an adaptive region clustering approach based on EM. We build the cookbook from the color, texture and SIFT features and perform the LDA training using Gibbs Sampling for topics. Then the adaptive region clustering with EM is invented to merge the regions based on topic distribution. The clustering number is self-identified according to Minimum Description Length (MDL) principle. And an image is represented as a Gaussian Mixture Model (GMM) with objects corresponding to Gaussian mixture components. The final segmentation could be achieved after the region clustering and adjacent check. We implemented the new algorithm and conducted experiments to validate the region clustering approach and segmentation performance. And the results show great effectiveness of this new algorithm.

---

X. Wang (✉) · J. Du · S. Wu  
Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia,  
Beijing University of Posts and Telecommunications, Beijing 100876, China  
e-mail: wxr@bupt.edu.cn

F. Li  
Department of Electrical and Computer Engineering, Portland State University,  
Portland, OR, USA

**Keywords** Topic model · LDA · Region clustering · Adaptive clustering · Image segmentation

## 66.1 Introduction

Image segmentation serves as a fundamental task for further explorations in image processing and understanding such as object recognition, image annotation, computer vision and etc. The task of segmentation is to partition an image into disjoint and semantically coherent regions, which generally forms meaningful objects [1]. Such perceptual grouping is more than just dividing an image based on low-level cues but further exploits higher level information or structure so as to produce a complete and correct segmentation [2]. The information can be co-occurring relationships, spatial semantics or in some other form.

Inspired by the effectiveness of Bag of Words (BOW) model in text analysis, researchers have introduced it into the field of image processing. However, different from texts, there are no real words in images. The general solution is to map the low-level image features, which are usually represented by feature vectors, into visual words. Cao and Fei-Fei [3] used SIFT descriptors for detected interest points as visual words and a corresponding codebook is obtained via k-Means algorithm.

BOW assumes that different words are independent of each other, and thus ignores the semantic connections between them, which are essential characteristics for both text and image. To catch the semantic connections between visual words, the aspect models such as LDA (Latent Dirichlet Allocation) [1] and pLSA (Probabilistic Latent Semantic Analysis) [4] are applied to vision processing tasks, such as image and region classification [5, 6], object recognition and segmentation [3, 7–9]. The aspect model is a powerful tool for modeling the co-occurring relationships between visual words based on low-level features, and such relationships are expressed by topics in the form of word-topic distributions. With such higher level semantics as topics, the performance of algorithms for vision analysis like image segmentation can be significantly improved.

Cao and Fei-Fei [3] established a Spatial Latent Topic Model based on LDA for image segmentation (Spatial LDA). It started with an over-segmentation of partitioning the image into small regions and then extracted the visual features of color, texture and SIFT descriptors for each over-segmented region, resulting in a visual words histogram for each region. These visual features were used for topic modeling in LDA. Consequently the regions were represented with topic distributions. The regions belonging to the same object would have similar topic distributions, thus in the same category. The Spatial LDA picked the topic with highest probability as the label for each region. If two regions had the same label and were also adjacent, they were merged as one.

However, there are two challenges for these image segmentation algorithms based on LDA idea. The first one is to determine the size of codebook and the number of topics. Different from the text analysis, the visual words obtained from clustering are a kind of statistical representation. Because of the high dimensions of feature vectors, a small codebook could cluster the feature vectors belonging to different objects into one word, while a large one risks separating similar feature vectors into different visual words. The similar problem also exists for the number of topics in LDA. The more topics are trained, more scattered the visual words will be. Thus it fails to catch the semantic connections between them so some topics tend to have very similar distributions. On the other hand, mixing different co-occurring relationships will weaken each one of them and also decrease the distinctions between topics. Both occasions will reduce the topic accuracy level and consequently degrade the final segmentation, especially when just a single topic with the highest probability is considered. The topic modeling result in LDA will be greatly impacted by these decisions so a single topic with the highest probability may not accurately represent the real semantics of the region. Instead, we propose to use the topic distribution for each region. The accuracy issue induced by the size of codebook and the topic numbers could be effectively reduced.

The second challenge is to determine the clustering number of clustering result. We use the clustering algorithm to automatically learn the topic categories according to topic distributions. Commonly used algorithms include k-Means, Fuzzy C-Means and etc., in which the clustering number needs to be specified in advance. In image segmentation, however, it is barely possible since there may have different number of objects in images.

According to the analyses above, we propose, in this paper, an adaptive image segmentation approach based on clustering with Expectation–Maximization (EM) algorithm. The topic category for each region is more accurately predicted with overall considerations for its topic distributions. And in order to adaptively identify the clustering number, Minimum Description Length (MDL) principle is adopted in EM instead of Maximum Likelihood (ML) principle, which can obtain an estimation of clustering number as close as possible to the real object number in the image.

We believe we have three innovative contributions in this paper. The first one is that LDA topic model is used for image regions so that the regions are represented by topic distributions, which have better high-level semantic connections. The second one is that we consider the topic distributions for region merging instead of single topic in previous papers. The final regions could have better correspondence with the high-level objects. The third contribution of this paper is that an adaptive clustering algorithm based on MDL is proposed to determine the clustering number adaptively, which is much appropriate for image analysis.

The rest of this paper is organized as following. The proposed approach will be illustrated in [Sect. 66.2](#) and the training of LDA with Gibbs Sampling is also described. Then the essential phase of adaptive region clustering is explained in details in [Sect. 66.3](#). [Section 66.4](#) presents the experiment data sets, and experiment results along with the analysis. The final section states the conclusions.



## 66.2 Segmentation Framework and LDA with Gibbs Sampling

### 66.2.1 Segmentation Framework

The proposed segmentation approach can be divided into four phases as illustrated in Fig. 66.1 and each of them is described in the following.

Phase 1: Over-segmentation and feature extraction. To start with, the over-segmentation of an image is obtained by using Normalized Cut [2], which will partition the image into reasonably small regions to ensure that no regions appear larger than the objects to be segmented. We also avoid too small regions, which will increase the model complexity and reduce the accuracy for the final segmentation. After the over-segmentation, three kinds of low-level features for each region are extracted: color, texture and SIFT. The color and texture are the regional appearance and SIFT features describe the detected interest points within the region [10, 11], which are in consistent with [3]. More specifically, the color and texture features are calculated by averaging over all the pixels within the region. And 2000-dimension SIFT features are chosen for interest points. Two codebooks are obtained correspondingly and the features are represented with indices in the codebooks.

Phase 2: LDA training with Gibbs Sampling. With images represented by regional appearances and visual words, LDA is then trained with Gibbs Sampling to model the co-occurrence relationships between visual features as latent topics. According to the graphical model of Spatial LDA (Fig. 66.2), a region is considered as the basic unit for topic assignment,

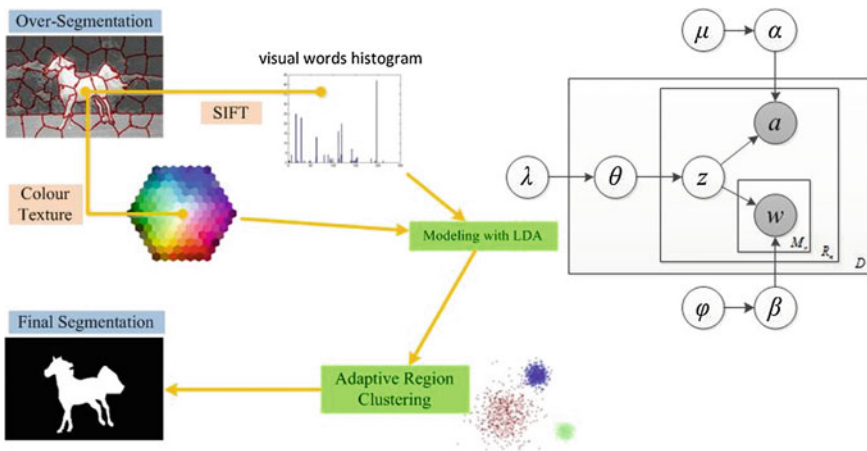


Fig. 66.1 Four phases of the proposed approach

enforcing semantic coherency within each one [3]. Topic-image distribution and word-topic distributions are estimated after training since they are first passed by Gibbs Sampling. More details of the training process with Gibbs Sampling are explained in Sect. 66.2.2.

- Phase 3: Adaptive region clustering with EM. With the estimated word-topic distributions, topic distributions are calculated for each over-segmented region in a given image and are used for clustering with EM. Our clustering approach adaptively estimates the number of clusters that best fit the data. It is based on the Minimum Description Length (MDL) principle by Rissanen [12]. The clustering can be interpreted with image representation of Gaussian Mixture Model (GMM), with each object as a Gaussian mixture component of the image. Details of this phase will be illustrated in Sect. 66.3.
- Phase 4: Segmentation. After the region clustering, each of the over-segmented regions has been identified as in some cluster in terms of their topic distributions. Each cluster corresponds to a category of regional topics, i.e. an object in the image. So each region is examined and it will be merged the other one if they are both clustered together and adjacent with each other. The final segmentation is formed after the merge of regions based on the clustering.

### 66.2.2 LDA with Gibbs Sampling

As for probabilistic inference in LDA, Gibbs Sampling is adopted in this paper. Gibbs Sampling is an effective approach to estimate joint distribution through conditional distribution. Each sampling conditions on the current state of the rest of the variables and the process is repeated until convergence [13–15].

With the addition of Dirichlet priors  $\mu$  and  $\varphi$  on  $\alpha$  and  $\beta$  respectively as in [14] for  $\phi$  (Fig. 66.2), the conditional distribution can be obtained as

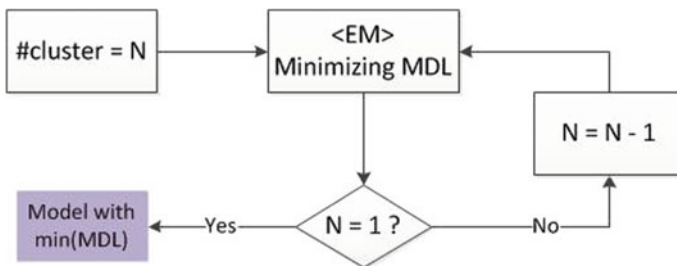


Fig. 66.2 Model selection with MDL

$$p(z_r = j | z_{-r}, \mathbf{a}, \mathbf{w}) = \frac{n_{j,-r}^{(d)} + \lambda}{n_{,-r}^{(d)} + T\lambda} \cdot \frac{n_{a_r,-r}^{(j)} + \mu}{n_{,-r}^{(j,a)} + A\mu} \cdot \frac{\prod_{k=1}^{V_r} \prod_{i=1}^{c_k} (n_{w_k}^{(j,w)} - t + \varphi)}{\prod_{k=1}^{M_r} \left[ \sum_{i=1}^w n_{w_k}^{(j,w)} + W\varphi - K \right]} \quad (66.1)$$

which is inducted similarly as in [14]. And there holds

$$\sum_{k=1}^{V_r} c_k = M_r \quad (66.2)$$

The notations are explained in Table 66.1.

Gibbs Sampling passes the topic-image distribution and word-topic distributions during LDA training, so the three distributions need to be estimated separately.

$$\begin{aligned} \theta_j^{(d)} &= \frac{n_j^{(d)} + \lambda}{n^{(d)} + T\lambda} \\ \alpha_k^{(j)} &= \frac{n_{a_k}^j + \mu}{n^{(j,a)} + A\mu} \\ \beta_k^{(j)} &= \frac{n_{w_k}^{(j)} + \varphi}{n^{(j,w)} + W\varphi} \end{aligned} \quad (66.3)$$

Since Gibbs Sampling is not the focus of this paper, we don't discuss further and readers can refer to [14] for more details of Gibbs Sampling for LDA and other aspect models.

**Table 66.1** Notations in formula (66.1) and (66.2)

Notation	Description
T	The number of topics trained with LDA
A	The size of codebook for region appearances
W	The size of codebook for visual words (SIFT descriptors)
$M_r$	The total number of visual words in region r
$V_r$	The number of distinct visual words in region
$C_k$	The appearance count in the current region of the kth visual word among the $V_r$ distinct ones
$n_{j,-r}^{(d)}$	The number of times topic j appears in image d, either in form of a regional appearance feature or a visual word (the dot in the subscript of $n^{(d)}$ means for all topics)
$n_{a_r,-r}^{(j)}$	The number of times that topic j is assigned to the appearance feature $a_r$ (the dot in the subscript of $n^{(j,a)}$ means for all appearance features, and note that the 'a' in the superscript indicates appearance features while a 'w' would indicate visual words)

## 66.3 Adaptive Regions Clustering with EM

After topic modeling with LDA, the co-occurrence relationships between visual features are represented by latent topics in the form of probability distribution, thus connecting low-level cues with higher level semantics. Topic distributions of each over-segmented region can then be calculated and be leveraged for the final merging process. In [3], Eq. (66.4) is used to label each region with the most likely topic.

$$z^* = \arg \max_z p(a, w|z) \quad (66.4)$$

As stated in Sect. 66.1, such label assignment ignores the overall distribution of topics within the region and is easily impacted by codebook and topic accuracy.

The previous analysis led to the introduction of clustering method for the identification of regional topic categories and then the problem of clustering number showed up. To address it, this paper adopts EM algorithm with MDL principle to adaptively cluster regions based on image representation with GMM.

### 66.3.1 Image Representation with GMM

An image can be viewed as a set of objects and each object consists of several regions. The relationship of regions composing an object can be described with Gaussian distribution and thus the image becomes a Gaussian Mixture Model (GMM). For example, assuming there is an image with three objects (e.g. sky, grassland and a horse) and it is partitioned into 50 regions. Then the image is a mixture of three objects according to the proportions of their areas, and on the other hand, each of the three objects is made up of some regions from the total 50 regions, which satisfies Gaussian distributions in the feature space.

With such representation, there are three model parameters to be estimated.

$\pi_k$ : the mixture proportions of cluster k in the model

$\mu_k$ : the mean vector of cluster k

$R_k$ : the covariance matrix of cluster k

And the probability of a region  $r_i$  belonging to a certain cluster k can be written as

$$p(r_i|\pi, \mu_k, R_k) = \frac{1}{(2\pi)^{M/2} |R_k|^{1/2}} \exp \left\{ -\frac{1}{2} (r_i - \mu_k)^T R_k^{-1} (r_i - \mu_k) \right\} \quad (66.5)$$

### 66.3.2 Address the Clustering Number Issue with MDL

To obtain an optimal segmentation, the number of clusters should be as close as possible to the real number of objects in the image. This paper adopts Minimum Description Length (MDL) principle [12] to adaptively identify the number of

clusters. Compared with the commonly used Maximum Likelihood (ML) principle, MDL can not only estimate the model parameters but also the number of parameters or even the structure of models. The idea behind MDL is to obtain a shortest description of both the data and the model, balancing between the fitness of the data and the complexity of the model. It can effectively avoid over-fitting issue. Moreover, MDL does not require there being a right-answer model but actually it seeks one with the optimal performance to predict unknown data [16].

In the problem of region clustering, the objective is to minimize the value of MDL which is calculated with Eq. (66.6) as in [17].

$$MDL(K, \pi, \mu, R) = - \sum_{i=1}^N \left( \sum_{j=1}^k \pi_j p(r_i | \pi, \mu_j, R_j) \right) + \frac{1}{2} L \log(NT) \quad (66.6)$$

$$L = K \left( 1 + T + \frac{T(T+1)}{2} - 1 \right)$$

where  $N$  is the number of regions in the image, and  $T$  is the number of topics trained with LDA. Bouman [17] also provides further and more detailed discussions on MDL usage with EM.

An iterative approach is adopted to identify the optimal number of clusters with MDL, as illustrated in Fig. 66.2. An initial value is needed for the number of clusters. But it does not have to be accurate or satisfy any special conditions except that it should not be too small. The half of the number of over-segmented regions is good and of course you can use a different one, say, just the number of over-segmented regions.

The iteration starts with the initial set of clustering number and EM is used to minimize the MDL and estimates the corresponding model parameters. When EM finish, one is subtracted from the clustering number unless it is equal to one. New iteration then starts with the new set of clustering number. The iteration stops if all of the possible clustering number is attempted and the model with the minimum MDL value will be selected for the final segmentation.

### 66.3.3 Model Estimation with EM

Expectation–Maximization (EM) algorithm is a common approach for model estimation in the situation of incomplete data. Here what is unknown is the object, which is unobserved for machines just as latent topics in LDA. The algorithm consists of the following two steps.

- E-step. The posterior probabilities of each cluster are calculated with current set of model parameters. Applying the Bayesian Rule, it can be obtained that

$$p(K|r_i, \pi, \mu, R) = \frac{\pi_k p(r_i|\pi, \mu_k, R_k)}{\sum_{i=1}^K \pi_i p(r_i|\pi, \mu_i, R_i)} \quad (66.7)$$

- M-step. With the newly calculated probability, model parameters are updated to minimize the value of MDL. (Refer to [17] for detailed inductions.)

$$\begin{aligned} \pi_k &= \frac{s}{N} \\ \mu_k &= \frac{1}{s} \sum_{i=1}^N r_i p(k|r_i, \pi, \mu, R) \\ R_k &= \frac{1}{s} \sum_{i=1}^N (r_i - \mu_k)(r_i - \mu_k)^T p(k|r_i, \pi, \mu, R) \\ s &= \sum_{i=1}^N p(k|r_i, \pi, \mu, R) \end{aligned} \quad (66.8)$$

The iteration stops with the convergence of MDL and the model parameters for current number of clusters are obtained.

## 66.4 Experiment

To validate the proposed approach, experiments are conducted for real segmentation and the clustering performance is also analyzed.

### 66.4.1 Data Set

To test the result of the proposed approach, the Microsoft object recognition data set [18] is used, which consists of 182 images of different cows, houses, trees and people with varied poses, colors and positions. The size of the images is  $426 \times 320$  and they are downsized to be no larger than  $320 \times 320$  for processing convenience. 100 images in the set are used for LDA training and the rest are for segmentations.

**Table 66.2** Best clustering numbers with k-Means and the ones chosen by EM for 6 images

Images	1	2	3	4	5	6
k-Means	5	7	7	6	7	6
EM	5	6	5	7	7	5

### 66.4.2 Region Clustering Results

Comparisons with traditional method such as k-Means are performed to validate the proposed region clustering approach based on EM. The two methods are applied to the same image and k-Means is experimented with a series of k values. Table 66.2 shows the best clustering numbers with k-Means for 6 images together with the one obtained with EM. The overall performance of EM is good while the results are little different from our anticipation, because the best clustering number is generally not the real number of objects in the images but larger than that. But it should be still normal since some of the visual details can hardly be noticed with bare eyes while the visual features extracted by machines will catch them and then be “seen” by machines.

### 66.4.3 Segmentation Results

With our algorithm, the segmentation accuracy was 75 % compared with 63 % for the Spatial LDA [3].

To validate the clustering performance of the proposed approach based on EM, the segmentation is conducted on some images, and some of the segmentation results are presented in Fig. 66.3.

**Fig. 66.3** Segmentation results

From the segmentation results, there can be seen clear profiles of the cows, trees, houses, people and the majority of the background is also merged, which indicates the effectiveness of the clustering approach with overall considerations for all topics trained with LDA. Some of the small pieces left unmerged come from the different visual effect with different light conditions as well as the accuracy of visual features. In addition, the number of images used for LDA training is a little bit small because of the limited image data sets, while a 2000-dimension full histogram of SIFT descriptor is used in our experiment as visual words. The small area regions, however, can be easily dealt with to be merged with the most similar large region around it.

## 66.5 Conclusions

In this paper, we seek to improve the LDA framework for image segmentation with an adaptive region clustering using EM algorithm. The proposed clustering approach makes thorough use of the topic information modeled by LDA and shows good performance. Moreover our clustering method can also be adopted in other aspect models for image segmentation.

**Acknowledgments** This work is supported by National Basic Research Program of China (973 Program) under Grant 2012CB821206, National Natural Science Foundation (NNSF) of China under Grant No. 91024001 and No.61070142, Beijing Natural Science Foundation under Grant No. 4111002, the Fundamental Research Funds for the Central Universities under Grant No.2013RC0306.

## References

1. Zhao B, Fei-Fei L, Xing E (2010) Image segmentation with topic random field. In: Proceedings of the 11th European conference on computer vision, pp 785–798
2. Shi J, Malik J (2000) Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Mach Intell* 22(8):888–905
3. Cao L, Fei-Fei L (2007) Spatially coherent latent topic model for concurrent segmentation and classification of objects and scenes. In: Proceedings of IEEE international conference on computer vision, pp 1–8
4. Hofmann T (2001) Unsupervised learning by probabilistic latent semantic analysis. *Mach Learn* 42(1):177–196
5. Fei-Fei L, Perona P (2005) A Bayesian hierarchical model for learning natural scene categories. In: Proceedings of IEEE computer society conference on computer vision and pattern recognition, pp 524–531
6. Verbeek J, Triggs B (2007) Region classification with markov field aspect models. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1–8
7. Russell BC, Freeman WT, Efros AA, Sivic J, Zisserman A (2006) Using multiple segmentations to discover objects and their extent in image collections. In: Proceedings of



- IEEE computer society conference on computer vision and pattern recognition, pp 1605–1614
8. Sivic J, Russell BC, Efros AA, Zisserman A, Freeman WT (2005) Discovering objects and their location in images. In: Proceedings of 12th IEEE international conference on computer vision, pp 370–377
  9. Wang X, Grimson E (2007) Spatial latent dirichlet allocation. *Adv Neural Inf Process Syst* 20:1577–1584
  10. Kadir T, Brady M (2001) Saliency, scale and image description. *Int J Comput Vis* 45(2):83–105
  11. Lowe DG (1999) Object recognition from local scale-invariant features. In: Proceedings of the 17th IEEE international conference on computer vision, pp 1150–1157
  12. Rissanen J (1983) A universal prior for integers and estimation by minimum description length. *Ann Stat* 11(2):416–431
  13. Casella G, George EI (1992) Explaining the Gibbs sampler. *Am Stat* 46(3):167–174
  14. Griffiths TL, Steyvers M (2004) Finding scientific topics. *Proc Nat Acad Sci USA* 101(1):5228–5235
  15. MacKay DJC (2003) Information theory, inference and learning algorithms. Cambridge university press, New York (Chapter 3)
  16. Griinwald PD, Myung IJ, Pitt MA (eds) A tutorial introduction to the minimum description length principle. *Advances in Minimum Description Length: Theory and Applications*. MIT Press, Cambridge, pp 3–81
  17. Bouman CA (2000) Cluster: an unsupervised algorithm for modeling Gaussian mixtures. <http://www.ece.purdue.edu/~bouman>
  18. Winn J, Criminisi A, Minka T (2005) Object categorization by learned universal visual dictionary. In: Proceedings of 10th IEEE international conference on computer vision, pp 1800–1807
  19. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J Mach Learn Res* 3:993–1022

# Chapter 67

## Methods of Recognizing True and Fake Smiles by Using AU6 and AU12 in a Holistic Way

Pingping Wu, Wenmin Wang and Hong Liu

**Abstract** Smile is one of the simplest forms of expressions that it is easy to recognize for human beings. It will be one of the most natural, straightforward and friendly ways in Human Computer Interaction (HRI) if a computer could catch the subtle expression, understand the inner state of human and meanwhile give its feedback according to the corresponding instance. In this paper, some different methods are proposed, to realize the recognition of true and fake smiles, based on facial action units from the research field of psychology and human behavior. In all of the methods we used, AU6 and AU12 are dealt with together in each example, which is different from AU recognition. Some popular feature extraction and classification methods such as Gabor wavelets, 2DPCA, Adaboost and SVM are used in the holistic way to implement the recognition. Images in our database are all frontal facial images with smiles of different types and levels from subjects of different countries with different colors and ages. Lots of experiments show that the best accuracy of our methods in recognizing true and fake smiles is close to 86 %, while people's true-fake-smile recognition ability is much lower.

**Keywords** Smile detection · Gabor wavelets · Facial action units · 2DPCA · SVM

### 67.1 Introduction

Smile is one of the most common facial behaviors in our daily life. It plays an important role in face to face interaction which is a human-specific direct and naturally preeminent way of communication. People smile out of various reasons

---

P. Wu (✉) · W. Wang · H. Liu  
Engineering Lab on Intelligent Perception for Internet of Things (ELIP),  
Peking University Shenzhen Graduate School, Room 105 Building G, Nanshan district,  
Shenzhen 518055, China  
e-mail: wangwm@pkusz.edu.cn

such as to be polite, to express his/her inside feelings or even to conceal his/her real feelings, which cause different types of smiles. From researches of Frank and Ekman [1], however, only one particular type of smile called the enjoyment smile accompanies experienced positive emotions such as happiness, pleasure, or enjoyment. Here, the enjoyment smile is defined as true smile and other types as fake ones.

In the last decades, people have done a lot of research on face detection, face recognition and facial expression analysis. Face detection is a rather important and prerequisite step because we can't get the face recognized or the facial expression analyzed before the face is detected. For recent years, researchers have proposed different kinds of methods and some of them achieve fairly good results. A boosted cascade of Haar features was proposed by Viola and Jones [2] to get the face detected and their system was very robust and had the fastest detection speed. After all, face detection approaches are either based on a holistic way or an analytic way [3]. In the holistic way, the face is regarded as a whole unit. In the analytic way, the face is detected by analyzing some important facial features first (e.g., the eyes and the lips). The overall location of the face is then determined by the location of the features in correspondence with each other.

The research on true and fake smile recognition is closely related to facial expression recognition. When it comes to the facial expression analysis, it is necessary to mention the Facial Action Coding System (FACS). Due to the richness and complexity of facial expression, behavior scientists realize that it is necessary to create an objective coding standard. FACS is the most objective and widely used method for measuring and describing facial behaviors. Most automatic facial expression recognition systems were studied with posed expressions. Nowadays, some studies transfer their attention and focus on spontaneous facial expression which involved muscles and dynamics were different from those posed ones. Lv and Wang [4] used head motion and AAM features to realize a spontaneous facial expression recognition. A survey of affect recognition methods for spontaneous expressions has been done by Zeng et al. [5]. A new data set named GENKI was collected by Whitehill et al. [6] which consists of 63,000 images for practical smile detection.

Nakano et al. [7] designed a true smile recognition system using neural networks, in which they didn't give an explicit description about the true smile mentioned in their paper and also didn't give a reason why a smile was a true one or not. If the result of the neural network could be regarded as the classification of true and false smiles, it could also be considered as a cluster of different smile intensities. Zhang et al. did a deceit-detection [8] in facial expressions in which the enjoyment expression was also involved. In order to carry out the detection, they used DBF (distance based features) and TBF (texture based features) corresponding to MCs and got the accuracy of 73.16 % in deceit detection in enjoyment. Hoque et al. explored temporal patterns to distinguish delight smiles from frustrated smiles. The best classifier distinguished between the patterns of spontaneous smiles under delighted and frustrated stimuli with 92 % accuracy. However, their work is based on video sequences in which the sound information is

also added. Differently, we put our hands to static images and have no dynamic information used. The deceit detection in posed smile and spontaneous smile has been done.

In this paper, we aim to find an automatic true/fake smile recognition method. As the dynamic information couldn't be derived from a static image, there is not enough information to differentiate the true and fake smile, which increases the difficulty of recognition. In this paper, we treat AU6 and AU12 together in each example to realize the recognition of true and fake smiles. Here, we define that a smile is a true one only if AU6 and AU12 both happen in a static smile image. The methods for recognition presented in our paper are robust for they could work with different races and ages and could tolerate the face off front in some extent. In addition, sufficient theoretical foundations are given from the viewpoint of psychology to explain why the true smile is different from the fake one and how to distinguish them.

The rest of this paper is organized as follows: Sect. 67.2 describes how to extract the features from the true and fake smiles images. In Sect. 67.3, the experimental results are analyzed, and the conclusions are drawn in Sect. 67.4.

## 67.2 Feature Extraction and Classification

### 67.2.1 Feature Representation

Gabor wavelets were widely used in image processing, pattern recognition and other fields due to their biological relevance and computational properties. Gabor filters are robust since it has the ability to hold with the rotation and deformation of images in some degree. For this advantage, the accuracy of our experiments is ensured because some smile images in our database are slightly yawed, pitched or rolled. Here, five different scales  $v \in \{0, \dots, 4\}$  and eight different orientations  $u \in \{1, \dots, 8\}$  are chosen to realize the Gabor filter. The image  $I(x, y)$  is convolved with the 40 Gabor kernels  $g_{\mu, v}(z)$  separately (five scales  $\times$  eight orientations),

$$W_{u, v}(x, y) = I(x, y) * g_{u, v}(x, y) \quad (67.1)$$

The magnitude response  $||W_{u, v}(x, y)||$  is used to represent the feature.

After Gabor filtering, the dimension is increased by 40 times. Using them as the feature directly will lead to high computational complexity and memory requirements. Furthermore, it is difficult to analyze such high-dimensional data accurately. In the following part, the dimension reduction is done by 2DPCA and Adaboost.

### 67.2.2 Feature Extraction Using 2DPCA

Principal component analysis (PCA) is probably one of the most popular techniques used for dimension-reducing. Yang et al. proposed the 2D-PCA approach whose basic idea is directly using 2D matrices to construct the corresponding covariance matrix instead of a 1D vector set, which improves the computational efficiency. The projection of a sample on each principal orthogonal vector is a vector and the problem of over-compression is alleviated in the 2D-PCA case. As outputs of Gabor filters are 2D matrices, it is more suitable to apply 2DPCA directly on them than PCA.

The convolution output of each Gabor filter contains different local, scale and orientation features. Instead of transforming each 2D convolution output into a vector, it makes sense to operate on them directly. Different from PCA, the covariance matrix  $C$  is defined as follows [9]:

$$C = \frac{1}{N} \sum_{k=1}^N (x_k - \bar{x})(x_k - \bar{x})^T \quad (67.2)$$

where  $\bar{x} = 1/N \sum_{k=1}^N x_k$  is the mean of the total training samples.

According to the generalized total scatter criterion:

$$J(v) = v^T C v \quad (67.3)$$

where  $v$  is a unitary column vector. The unitary vector  $v$  that maximizes the criterion is called the optimal projection axis. Intuitively, this means that the total scatter of the projected samples is maximized after the projection of an image matrix onto  $v$ .

$$\begin{cases} \{v_1, \dots, v_d\} = \arg \max J(v) \\ v_i^T v_j = 0, \quad i \neq j, \quad j = 1, \dots, d. \end{cases} \quad (67.4)$$

In fact, the optimal projection,  $v_1, \dots, v_d$  are orthonormal eigenvectors of  $C$  corresponding to the first  $d$  largest eigenvalues. For any output of Gabor filters  $W$ , its projection to this group of optimal projection vectors is:

$$y_i = W v_i, \quad i = 1, \dots, d \quad (67.5)$$

Therefore,  $y_1, \dots, y_d$  are called the principal component (vector) of the sample image.

### 67.2.3 Feature Extraction Using Adaboost

Adaboost is not only a fast classifier but also an effective feature selection method. The basic idea of Adaboost algorithm is that a strong classifier could be expressed

as the linear combination of a series of weak classifiers with different weights on the training set. Here Adaboost is used to extract features and treat each Gabor filter as a weak classifier. Adaboost picks up the best one of these classifiers and boosts the weights on the error examples. The next filter is selected which gives best performance on the errors of the previous one. After  $T$  rounds of iteration,  $T$  features are selected out. The weak classifier could be expressed as follows:

$$h_j(x) = \begin{cases} +1, & p_j\phi_j(x) < p_j\theta_j \\ -1, & \text{otherwise} \end{cases} \quad (67.6)$$

where  $x$  is an example,  $\phi_j(x)$  represents extracting a feature from  $x$ , and  $p_j$  is the sign which maintains the direction of the inequality. The detailed steps of attribute sorting with Adaboost algorithm could be seen in [10].

### 67.2.4 Feature Classification

Adaboost, SVM, LDA and BP Network are all familiar classifiers. Both SVM and Adaboost could deal with high dimensional space and are simple to train and perform in real time. Their generalization ability is well. A deep inside analysis has been done in [11] to narrate the similarities and differences between them. Special examples (support vectors) are selected by SVM while particular features are selected by Adaboost. It is very important to know that Adaboost is not only a fast classifier but also an effective feature selection method. SVM has been shown to perform better when the feature space is dense which means the features are highly relevant to each other [12]. Experiments have been done in [11] to explore training SVM with the features selected by Adaboost. And the results show that training SVM on the continuous outputs of the selected filters of Adaboost outperforms Adaboost and SVM individually. Linear Discriminant Analysis (LDA) has been shown to be also widely used in facial expression recognition in many works. LDA is more suitable to classify the examples of Gaussian distribution while SVM not. BP network is based on empirical risk minimization and easy to trap in local optimum while SVM is based on structure risk minimization and considers the sample error and the model complexity. Local optimum is global optimum in SVM. SVM shows better generalization ability than BP network. Here, SVM, LDA and BP Neural Network are chosen to complete the task of classification. Experiments are done in Sect. 67.3 to find out which classifier is best for the feature selected above in this specific mission.

The overall procedure of our true/fake smile recognition system is shown in Fig. 67.1, which will be stated in detail in the following section.

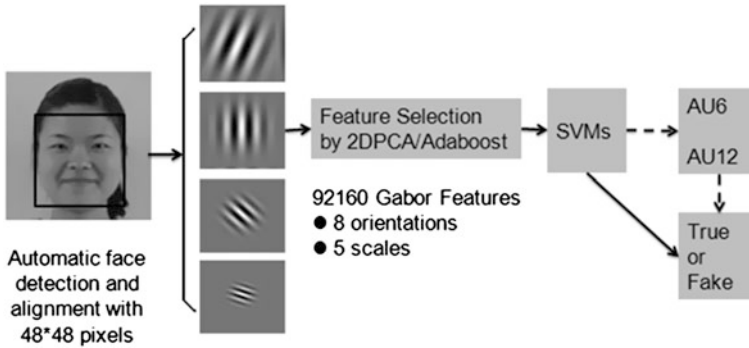


Fig. 67.1 Automated true and fake smile recognition system

## 67.3 Experiments and Analysis

Our experiments are implemented in Matlab and C++. LIBSVM from C. Lin is used and the linear kernel is chosen. We use the free source code of face detector available at [mplab.ucsd.edu](http://mplab.ucsd.edu) which is an improved version and has been shown to perform rather well.

### 67.3.1 Database

Images in database are all of smiling-face images from front view. 220 images of different kinds of smiles are gathered with half true and half fake. All the images are colored and saved with resolution of  $256 \times 256$ . 100 of them are captured from public database (BBC: Human Body and Mind) from 20 objects, seven females and 13 males with five pictures per person of different ages and races. The rest 120 are created by ourselves for 12 subjects whom are all Chinese aged from 20 to 25 with 10 pictures per person. In the 12 subjects, five of them are female and seven male. The 100 images collected from the public database have been analyzed and labeled by the author already, which is shown in Fig. 67.2. For the rest 120 images, some were taken when the subjects were watching some funny films and some of them were taken when the subject just posed a smile. We analyze and label them according to FACS. All the smile images are captured from the video sequences at the smile apexes manually. All the used subjects are healthy people without any disease of the facial muscle.

The Alignment is realized in the automatic way, in which automatic eye detection is used to find the centers of the two eyes and then the image is rotated to make the eyes horizontal. All the faces and eyes in our database have been successfully detected.

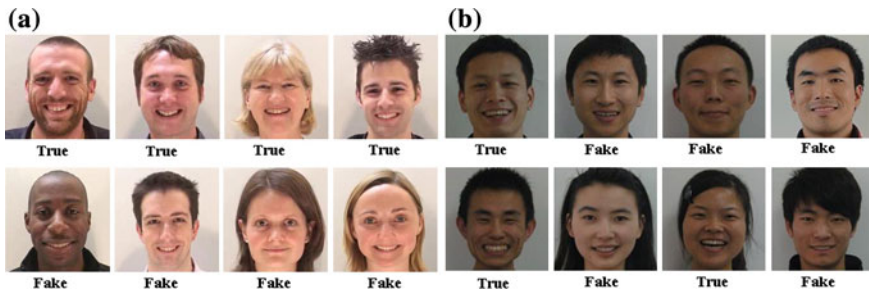


Fig. 67.2 a Part of the smile faces collected from BBC and b part of the smile faces captured by ourselves

### 67.3.2 2DPCA-SVM

After obtaining the face region and resizing it to  $48 \times 48$  pixels, the Gabor filter is applied. This will keep unchanged in other methods' implementation. Different number of principal components  $d_1 = d_2 = 1, 2, 4, 6, 8, 10, 12, 16, 20$  and  $24$  are tried as shown in Fig. 67.3. When  $d_1 = d_2 = 8$ , the hit rate is tending to reach the apex. Therefore,  $d_1 = d_2 = 8$  are used, which make the dimensionality reduced by 97.2 %. The total dimensionality of the extracted feature is  $40 \times 8 \times 8$ .

The 220 images are divided into five subsets of equal size randomly. Sequentially, one subset is tested using the classifier trained on the remaining four subsets. Therefore, each instance of the whole set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified. This procedure can prevent the over-fitting problem. Results of 2DPCA-SVM are shown in Table 67.1, from which it could be found that the method has better performance when recognizing fake smiles.

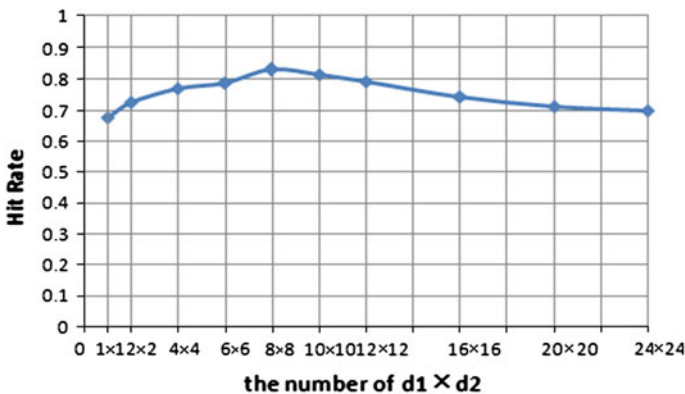


Fig. 67.3 Recognition accuracy with different number of  $d_1$  and  $d_2$



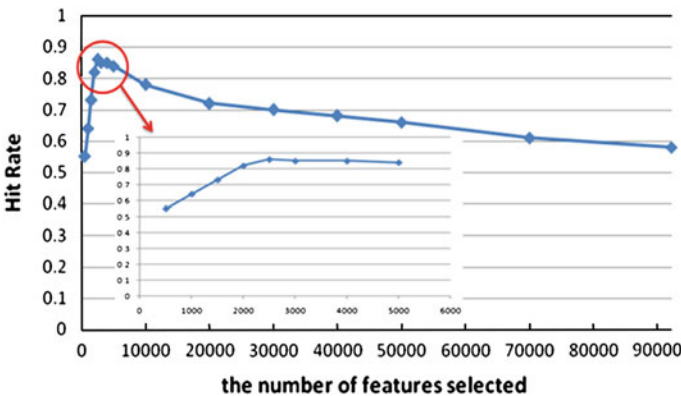
**Table 67.1** Results of 2DPCA-SVM and adaboost-SVM with cross-validation

	True positive rate	True negative rate	Hit rate
2DPCA-SVM	80 % (88/110)	83.6 % (92/110)	81.8 % (180/220)
Adaboost-SVM	86.4 % (95/110)	85.4 % (94/110)	85.9 % (189/220)

### 67.3.3 Adaboost-SVM

Instead of 2DPCA, the Adaboost algorithm introduced in Sect. 67.2.3 is tried to get better results. Different feature numbers are selected to find out the most proper amount of features. As shown in Fig. 67.4, the hit rate comes to its peak around 2,500. Therefore, 2,500 features are chosen as the input features of the classifiers. Results with cross validation are also shown in Table 67.1. It could be found that all of the rates have been improved through the method. In the meantime, the difference in TPR and TNR is smaller than 2DPCA-SVM. With Adaboost-SVM method, 189 smile pictures are correctly recognized while with 2DPCA-SVM method, 180 pictures are rightly classified. In the two correctly recognized sets, 168 pictures are in common. 81 true smile pictures classified by 2DPCA-SVM are also recognized as true with Adaboost-SVM while 87 fake smile pictures which are correctly recognized are the same in both methods. Therefore, if covering the true recognized pictures in both methods, the hit rate will be improved to 91.4 %.

For different feature selection methods and classifiers, we try to dig out which feature selection method is better and which classifier is more suitable. The best combination needs to be found out. Corresponding to 2DPCA, 2,560 PCs are selected by PCA. The BP network is of three layers and the input layer has 64 nodes. The output layer is of 2 nodes to represent a true or fake smile. Twelve nodes are used as the hidden layers. As shown in Table 67.3, results from the SVM classifier are better than the other two classifiers. One of the reasons is that the SVM classifier is more suitable for small sample data analysis. When using PCA to



**Fig. 67.4** Recognition accuracy for different number of features selected by Adaboost

**Table 67.2** Results of different feature selection methods with different classifiers with cross validation

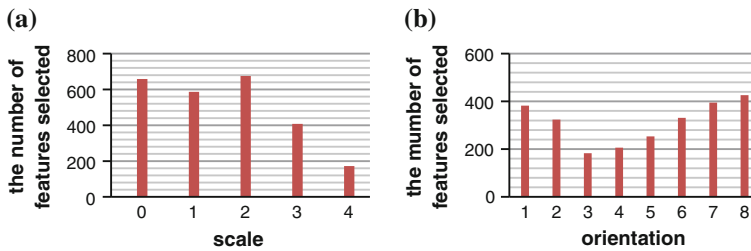
	LDA (%)	SVM (linear) (%)	BP network (%)
PCA	80.0	79.5	70.9
2DPCA	79.1	81.8	74.1
Adaboost	84.6	85.9	77.8

select the feature, LDA outperforms the other two. As 2DPCA is applied, SVM achieves the best result. When it comes to Adaboost, the three classifiers' performances are all improved and still SVM overcomes the other two. From the experiments, Adaboost is the best feature selection method and SVM the best classifier. In the meantime, from the data in Table 67.3, it is found that they are the best combination for the task (Table 67.2).

In the 2,500 features selected by Adaboost from the Gabor features, we try to learn the distribution of them in each channel which means different scales and orientations. Apparently, different channels of Gabor filters do different levels of contribution for the task. Fig. 67.5 shows the distributions of the extracted features in five scales and eight orientations. We could see that the scales  $v = 0, 1$  and  $2$  are more contributive than the other two. And orientations  $u = 1, 7$  and  $8$  are more dedicative than the rest.

### 67.3.4 Analysis

Experiments also have been done to check out the methods' ability in recognizing the true/fake smile of different races and genders. In the comparison of different races, the database is divided into two categories. The first category is 120 pictures with subjects all Asians and the second one is 100 pictures with subjects non Asians. When comparing two different genders, the database is divided into two parts with one part all female and the other all male. Table 67.3 shows the results (Hit Rate) from two methods of cross-validation with Gabor filters of five scales and eight orientations. It is found that both methods prefer Asians. Both methods' Hit Rate dropped when dealing with non Asians. One of the reasons leading to the phenomenon is that the coverage of the 120 pictures is not wide enough and the subjects' age concentrates in 20s. On the other hand, the other 100 pictures are from subjects with different ages, colors and races. The hit rate of female pictures and the hit rate of male pictures are quite similar with both two methods. The data shows that the recognition of true and fake smile is irrespective of gender to some extent.



**Fig. 67.5** Distributions of Gabor features **a** in different scales and **b** in different orientations

**Table 67.3** Comparison of performance of different races and genders through different methods with five-scale and eight-orientation gabor filters

	Adaboost-SVM	2DPCA-SVM
Asians (120 pictures)	86.7 % (104/120)	84.2 % (101/120)
Non Asians (100 pictures)	85.0 % (85/100)	79.0 % (79/100)
Mixed (220 pictures)	85.9 % (189/220)	81.8 % (180/220)
12 Females(85 pictures)	85.9 % (73/85)	82.3 % (70/85)
20 Males(135 pictures)	85.9 % (116/135)	81.5 % (110/135)

## 67.4 Conclusions

In this paper, we combine with the facial action units of FACS to analyze true and fake smiles. The recognition is realized in the holistic way by dealing with AU6 and AU12 together, which is different from AU recognition. Gabor filters are used to present features as they have excellent performance in texture representation and discrimination. In order to reduce the dimension to make the analysis and computation easier, 2DPCA and Adaboost are applied respectively. Finally, different classifiers such as SVM, LDA and BP neural network are used to recognize true and fake smiles. Comparison experimental results show that the best combination of methods is Adaboost + SVM.

From the experiments, it is found that the recognition of true and fake smiles is independent to the gender. The hit rate could still be improved by analyzing more detailed features. Anyway, the true-fake-smile recognition system would improve the human robot interaction and make the interaction more friendly and deeply. Furthermore, it could also be used as a tool for behavioral science and psychology research, which is worth studying.

**Acknowledgments** This work is supported by National Natural Science Foundation of China (NSFC, No. 60875050, 60675025), National High Technology Research and Development Program of China (863 Program, No. 2006AA04Z247), Scientific and Technical Innovation Commission of Shenzhen Municipality (No. JC201005280682A, CXC201104210010A).

## References

1. Frank M, Ekman P (1993) Not all smiles are created equal: the difference between enjoyment and nonenjoyment smiles. *Int J Humor Res* 6(1):9–26
2. Viola P, Jones M (2004) Robust real-time face detection. *Int J Comput Vis* 57(2):137–154
3. Pantic M, Rothkrantz L (2000) Automatic analysis of facial expressions: the state of the art. *IEEE Trans Pattern Anal Mach Intell* 22(12):1424–1444
4. Lv Y, Wang S (2010) A spontaneous facial expression recognition method using head motion and AAM features. In: 2010 Second world congress on nature and biologically inspired computing, Dec. 15–17, pp 334–339
5. Zeng Z, Pantic M, Roisman G, Huang T (2009) A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Trans Pattern Anal Mach Intell* 31(1): 39–58
6. Whitehill J, Littlewort G, Fasel I, Bartlett M (2009) Toward Practical Smile Detection. *IEEE Trans Pattern Anal Mach Intell* 31(11):2106–2111
7. Nakano M, Mituskura Y, Fukumi M, Akamatsu N (2002) True smile recognition system using neural networks. In: Proceedings of the 9th international conference on neural information processing, vol 2, pp 650–654
8. Zhang Zhi, Singh Vartika, Slowe Thomas E, Tulyakov Sergey, Govindaraju Venuopal (2007) Real-time automatic deceit detection from involuntary facial expressions. *IEEE Conf Comput Vis Pattern Recogn* 26:131–137
9. Yang J, Zhang D, Frangi A, Yang J (2004) Two dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Trans Pattern Anal Mach Intell* 26(1):131–137
10. Wu Y, Liu H, Zha H (2005) Modeling facial expression space for recognition. In: International conference on intelligent robots and system, pp 1814–1820
11. Littlewort G, Bartlett MS, Fasel I, Susskind J, Movellan J (2006) Dynamics of facial expression extracted automatically from video. *Comput Vis Pattern Recogn Workshop* 24:615–625
12. Roth D, Yang M-H, Ahuja N (2002) Learning to recognize three dimensional objects. *Neural Comput* 14:1071–1103

# Chapter 68

## Constrained Silhouette Based Evolutionary K-Means

Zhenfeng He

**Abstract** Evolutionary K-Means (EKM) is a non-parametric approach proposed to improve K-Means algorithm. Current EKM approaches are ineffective in deciding the correct cluster number of real datasets. This paper uses instance-level constraints to solve this problem and presents a Constrained Silhouette (CS) based algorithm, namely CS-EAC. Firstly CS is defined to combine constraints into the computation of Silhouette Information (SI). Updated from the Fast Evolutionary Algorithm for Clustering algorithm (F-EAC), CS-EAC uses CS instead of SI to guide the genetic operations. Experimental results suggest that CS-EAC is effective in both deciding the correct number of clusters and improving the accuracy of clustering for real datasets.

**Keywords** Clustering · Semi-supervised clustering · Genetic algorithm · Silhouette information

### 68.1 Introduction

Clustering is a data analysis technique that groups instances into several clusters. There are many clustering algorithms; K-Means may be the most popular one. But K-Means is ineffective when the parameters  $K$  (the number of clusters) and  $C_0$  (initial centers of clusters) are inappropriate [1]. So K-Means is often executed multiple times with different parameters to get different partitions. Then these partitions are evaluated to find the best one. Because there are too many possible combinations of  $K$  and  $C_0$ , this approach will be inefficient if these parameters are randomly selected. Evolutionary K-Means (EKM) is an active research area that considers past execution results to select the parameters intelligently [2].

---

Z. He (✉)

College of Mathematics and Computer Science, Fuzhou University, 2 Xue Yuan Road, University Town, Fuzhou, Fujian, China  
e-mail: hezhenfeng@fzu.edu.cn

EKM is developed from Genetic Algorithm (GA) based clustering algorithms. Pure GA-based clustering approaches are usually bothered by two problems: slow convergence and hard-to-interpret results. EKM can solve these problems by using K-Means to do local search quickly and using K centers to represent the clusters just as K-Means does [3–5].

There are two different EKM approaches. Some still require users to input K, while others do not. The latter is the topic of this paper. Deciding K is difficult. Except for a few unsuccessful attempts, the performance of EKM in deciding K has been evaluated mostly on artificial datasets [4–6]. So a novel constrained clustering approach will be introduced to improve EKM's ability of deciding K. Constrained clustering has been a hot research topic for the past decade [1, 7]. There have been some related researches on constrained GA-based clustering. For example, Cop-CGA used constraints in density estimation [8]. Cop-HGA used One-step Constrained K-Means (OCK) operator to solve the backtracking problem [9]. But both Cop-CGA and Cop-HGA were proposed for the K-fixed problem. There were also limited tries on deciding K with constraints, but they were not in the framework of EKM [10]. The research on constrained EKM approaches to decide K is still not seen.

This paper will introduce constraints into Silhouette Information (SI) to guide the clustering process. It will firstly introduce EKM and SI, then give a short introduction on constrained clustering. After that, Constrained Silhouette (CS) will be defined, and the performance of the CS based EKM will also be presented.

## 68.2 Evolutionary K-Means

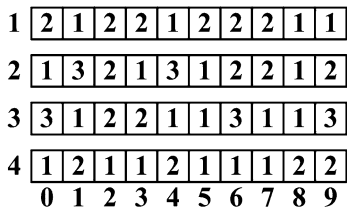
EKM is a hybrid approach of GA and K-Means [2, 4, 5]. Its framework is:

Input: dataset  $D$ , minimum/maximum cluster number  $KMIN/KMAX$

Output: partition  $C$ .

1. Initialization. Generate a population  $P$  randomly.
2. K-Means Process. For each gene  $g$  in  $P$ :
  - 2.1 Initialize a K-Means process with  $g$
  - 2.2 Do K-Means to get a partition  $p$
  - 2.3 Generate a new gene  $g_1$  from the partition  $p$  and replace  $g$  with  $g_1$
 New genes constitutes a new population  $P_1$ .
3. Genetic Algorithm Process (GA Process).
  - 3.0 Calculate the fitness value of all genes in the population  $P_1$ .
  - 3.1 Selection. Select genes from  $P_1$  to generate a new population  $P_2$ .
  - 3.2 Cross. Perform cross operation on  $P_2$  to generate a new population  $P_3$ .
  - 3.3 Mutation. Perform mutation operation on  $P_3$  to get a new population  $P$ . Adjust the parameters for the next iteration.
4. Go step 2 to evolve the new population if the stopping condition is not met.

**Fig. 68.1** A four-gene population for evolutionary K-means



EKM approaches often use genes to represent clustering results directly, that is, each gene is an ordered sequence of the labels. Figure 68.1 presents a four-gene population. The length of each gene is 10; so the dataset consists of 10 instances. The gene (2,1,2,2,1,2,2,2,1,1) defines a one-cluster partition. One cluster contains instances 1,4,8,9; another cluster contains the remaining instances.

A gene is used to initialize a K-Means Process of EKM (step 2.1). If gene (2,1,2,2,1,2,2,2,1,1) is used, then  $K = 2$ , and one cluster center is the mean of instances 1,4,8,9, another center is the mean of the remaining instances. Step 2.2 sometimes becomes a K-Means Operator (KMO), which means the assigning of instances and the updating of the centers are done only once [3].

The GA Process (step 3) is guided by the quality of the current partition. Silhouette Information (SI) is a popular validation criterion for EKM approaches [2, 6, 11]. This paper will be based on SI. Let  $dis(x, C)$  denote the distance between an instance  $x$  and a cluster  $C$ . When  $x$  belongs to  $C$ ,  $dis(x, C)$  will be denoted as  $a(x)$ . When  $C$  is the nearest of all the clusters which do not contain  $x$ ,  $dis(x, C)$  will be denoted as  $b(x)$ . The SI of an instance  $x$  is:

$$SI(x) = \frac{b(x) - a(x)}{\max\{b(x), a(x)\}} \tag{68.1}$$

Usually,  $dis(x, C)$  refers to the average distance between  $x$  and all instances in  $C$  (except of  $x$ ). So  $a(x)$  is the average difference between  $x$  and all other instances that grouped together with  $x$ ,  $b(x)$  is the average difference between  $x$  and all instances in the group which is the second closest. A large SI value for instance  $x$  means the second choice is not good, thus suggests  $x$  is “well clustered” [11].

The SI of a cluster or a partition can also be defined. A cluster’s SI is the average SI value of every instance in that cluster. A partition’s SI is the average SI value of all instances in that partition. A large SI value suggests that a cluster or a partition is good [11].

Different EKM approaches may have different GA Processes (step 3). This paper will use the GA Process of the Fast Evolutionary Algorithm for Clustering (F-EAC). F-EAC is fast because of its simple and efficient GA operators [2, 4, 5]. Without the rather inefficient Cross operation (step 3.2), F-EAC uses an elitist strategy in the Selection (step 3.1) and two adaptive mutation operators in the Mutation (step 3.3) [4, 5].

### 68.3 Constrained Clustering

Wagstaff et al. have defined instance level constraints to improve K-Means [1, 7]. They were Must-Link (ML) and Cannot-Link (CL) constraints. Instances  $x$  and  $y$  are Must-Linked, which will be denoted as  $ML(x,y)$ , if they should be grouped together; they are Cannot-Linked (will be denoted as  $CL(x,y)$ ) if they should NOT.

COP-Kmeans (CKM) is a popular instance-level constraint based K-Means algorithm [7]. It assigns instances based on their constrained distance  $ConDis$ . If the minimum constrained distance of an instance is  $+\infty$ , which means there is no feasible cluster for the instance, CKM will restart the clustering process.

$$ConDis(x, C) = \begin{cases} +\infty & \text{when } \exists y \in C, CL(x, y) \\ 0 & \text{when } \exists y \in C, ML(x, y) \\ d(x, O) & \text{otherwise} \end{cases} \quad (68.2)$$

where  $C$  is a cluster and  $O$  is its center. CKM tries to satisfy all constraints. This can lead to a high computation cost because CKM may need to restart too many times [8]. So it is impractical to include CKM in EKM.

One-step Constrained K-Means operator (OCK), which assigns instances and updates centers only once, appears to be an alternative [3, 9]. OCK also assigns instances based on  $ConDis$ ; but an instance will be added to a temporary buffer rather than restart the clustering when there is no feasible cluster for that instance. All instances in the buffer will be assigned based on unconstrained distance later. So backtracking is no longer needed. But OCK may be inefficient when  $K$  is unknown. Suppose the current cluster number is  $K_c$  and the real cluster number is  $K_r$ . When  $K_c < K_r$ , instances from different clusters must be crowded in a group, which makes  $CL$  difficult to be satisfied, and  $ML$  becomes easy to be satisfied at the same time. When  $K_c > K_r$ , we can have a similar analysis. So this paper will try to include the constraints in the GA Process instead.

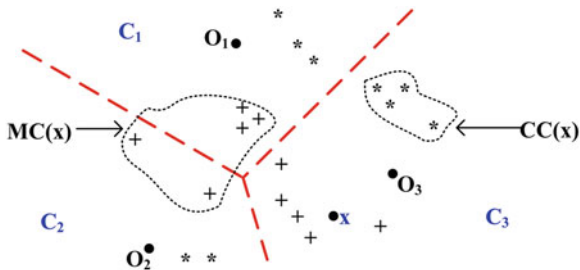
### 68.4 Combining Evolutionary K-Means with Constraints

Figure 68.2 is a partition of three clusters. Instance  $x$  belongs to cluster  $C_3$  (whose center is  $O_3$ ). Instances that have constraints with  $x$  are also presented in the figure. The instances that Must-Linked with  $x$  are denoted as '+', and those Cannot-Linked with  $x$  are denoted as '\*'.

We can define 2 sets for instance  $x$ , namely ML-Conflict set  $MC(x)$  and CL-Conflict set  $CC(x)$ .  $MC(x)$  consists the instances whose  $ML$  constraints with  $x$  are violated.  $CC(x)$  consists the instances whose  $CL$  constraints with  $x$  are violated. In Fig. 68.2,  $MC(x)$  is the set of instances which are not in  $C_3$  but are Must-Linked with  $x$ .  $CC(x)$  is the set of instances which are in  $C_3$  but are Cannot-Linked with  $x$ .



**Fig. 68.2** Must-Link (ML) conflict set and Cannot-Link (CL) conflict set of an instance  $x$



To make the figure clear and simple, the instances in  $MC(x)$  and  $CC(x)$  are closely distributed. In fact, they can be widely dispersed.

$CC(x)$  and  $MC(x)$  should be considered when computing  $SI(x)$ . The instances in  $CC(x)$  grouped together with  $x$  in the current partition. This implies that they are close to  $x$ . But they should not be grouped with  $x$  according to the constraints. So they may actually belong to the group which is the second closest to  $x$  in the user expected partition. This makes it necessary to consider them when computing  $b(x)$ . Following a similar analysis, we can conclude that the instances in  $MC(x)$  should be considered when computing  $a(x)$ . Therefore  $a(x)$  (when  $MC(x)$  is not null) and  $b(x)$  (when  $CC(x)$  is not null) can be redefined as:

$$a(x) = \max(dis(x, MC(x)), dis(x, C_m)) \tag{68.3}$$

$$b(x) = \min\left(dis(x, CC(x)), \min_{i \neq m} dis(x, C_i)\right) \tag{68.4}$$

where instance  $x$  belongs to cluster  $C_m$  in the current partition. When  $a(x)$  and  $b(x)$  are computed by (68.3) and (68.4), the  $SI$  value computed by formula (68.1) contains the constraint information, so it can be named as Constrained Silhouette (CS). It seems reasonable to define  $a(x)$  as Eq. (68.5) does. Yet, the instances belong to  $C_m$  may not be grouped together with  $x$  in the expected partition. It would be helpful to let  $MC(x)$  play a bigger role.

$$a(x) = dis(x, MC(x) \cup C_m) \tag{68.5}$$

The range of CS is  $[-1, 1]$ . To adjust the range to  $[0, 1]$ , CS is normalized by  $(1 + CS)/2$ . The normalized CS will be used in GA Process (step 3).

Each instance has the same weight when computing the  $SI$  of a cluster or a partition. As constraints are included, instances (clusters/partitions) that satisfy more constraints should be favored, so a weighted approach may be better. We can redefine the CS value of a cluster  $C$  or a partition  $P$  for a data set  $D$  as:

$$CS(C) = \frac{1}{|C|} \sum_{x \in C} w(x) CS(x) \tag{68.6}$$

$$CS(P) = \frac{1}{|D|} \sum_{x \in D} w(x) CS(x) \quad (68.7)$$

where  $|\cdot|$  stands for the cardinality of a set,  $w(x)$  is the weight for the instance  $x$ . If  $x$  has constraints,  $w(x)$  is the proportion of constraints on  $x$  that is satisfied. If  $x$  has no constraint,  $w(x)$  is the overall proportion of constraints satisfied by the cluster which contains  $x$ . The partitions and the clusters that have a higher constraint satisfaction rate are more likely to have a larger CS value. So they will have an advantage over other clusters or partitions in the evolutionary process.

Using CS to guide the GA Process of F-EAC (presented in [4, 5]), we can get a CS based EKM approach, which can be named as **CS-based EAC** (CS-EAC).

## 68.5 Experiments

Four datasets were used for evaluation. One artificial dataset “Data\_4\_3” was used, whose detailed description is in [4]. Three UCI datasets were also used. Table 68.1 gives a brief description of these datasets.

F-EAC [5], OCK-EKM [9] (EKM with OCK operator) and CS-EAC algorithm were tested. 100 randomly generated constraints were used to guide OCK-EKM and CS-EAC in the test. A transitive closure was taken to add the constraints just as CKM does [7]. A five time’s 10-fold cross validation was performed for each algorithm. Rand Index was used to measure the accuracy of clustering [7]. The algorithms stopped when there were no improvements during the past 10 iterations. The population size was 10. The results were given in Table 68.2.

Both the accuracy of clustering (outside parentheses) and the accuracy on deciding K (in parentheses) are presented in Table 68.2. CS-EAC’s performance was clearly the best. Compared with F-EAC, CS-EAC increased Iris’s clustering

**Table 68.1** Descriptions of the data sets

Dataset	Size	Feature	Class
Data_4_3	400	3	4
Iris	150	4	3
Glass	214	9	6
Column3C	310	6	3

**Table 68.2** The accuracy of clustering and the accuracy on deciding cluster numbers

Dataset	F-EAC	CS-EAC	OCK-EKM
Data_4_3	<b>100 % (100 %)</b>	<b>100 % (100 %)</b>	<b>100 % (100 %)</b>
Iris	76.6 % (0 %)	<b>82.3 % (56 %)</b>	76.8 % (4 %)
Glass	34.7 % (0 %)	<b>60.4 % (10 %)</b>	51.2 % (0 %)
Column3C	42.7 % (0 %)	<b>64.9 % (58 %)</b>	61.8 % (0 %)

accuracy from 76.6 to 82.3 %, at the same time it increased the accuracy on deciding K from 0 to 56 %. Even for glass data (which has 2 small clusters, each of which has only 10 instances), CS-EAC managed to increase the clustering accuracy from 34.7 to 60.4 %. This suggests that constraints can help EKM decide an appropriate K and improve clustering accuracy. OCK-EKM could increase the clustering accuracy, but it failed in deciding K.

The time cost of CS-EAC was about the same as that of F-EAC. When the clustering task was difficult, CS-EAC became slower because more iterations were needed to make constraints satisfied. For Glass data, CS-EAC run 15.9 generations on average, but F-EAC run 14.0 generations.

## 68.6 Conclusion

The initialization problem of K-Means often puzzles its users. EKM is a promising parametric-free framework proposed to solve this problem. But current EKM approaches are incapable of deciding K for many real data sets, which usually contain a lot of noise. Some background knowledge may be needed to improve EKM. This paper suggests using instance level constraints to guide the clustering process and presents a Constrained EKM approach. Constrained Silhouette (CS) is defined to handle the instances whose constraints are violated. To augment the influence of the constraints, a weighted approach for computing the CS of a cluster or a partition is given, which favors the partitions or clusters that satisfy more constraints. As a CS-based EKM approach, CS-EAC, which combines CS with F-EAC, is presented. Experimental results prove that CS is effective in help CS-EAC decide K for real data sets.

**Acknowledgments** This research was supported by Scientific Research Fund of Fujian Provincial Education Department (NO. JA11015).

## References

1. Jain AK (2010) Data clustering: 50 years beyond K-means. *Pattern Recogn Lett* 31(8):651–666
2. Hruschka ER, Campello RJGB, Freitas AA, Carvalho ACPLF (2009) A survey of evolutionary algorithms for clustering. *IEEE Trans Syst Man Cybern Part C Appl Rev* 39(2):133–155
3. Chen S, Chao Y, Wang H, Fu H (2006) A prototypes-embedded genetic K-Means algorithm. In: *Proceedings of the 18th international conference on pattern recognition (ICPR 06)*, pp 724–727
4. Alves V, Campello RJGB, Hruschka ER (2006) Towards a fast evolutionary algorithm for clustering. In: *Proceedings of the IEEE congress on evolutionary computation (CEC 06)*, pp 1776–1783

5. Naldi MC, Campello RJGB, Hruschka ER, Carvalho ACPLF (2011) Efficiency issues of evolutionary K-Means. *Appl Soft Comput* 11:1938–1952
6. Tseng M, Chiang C, Tang P, Wu H (2010) A study on cluster validity using intelligent evolutionary K-Means approach. In: *Proceedings of the 9th international conference on machine learning and cybernetics (ICMLC 10)*, pp 2510–2515
7. Wagstaff K, Cardie C, Rogers S, Schroedl S (2001) Constrained K-means clustering with background knowledge. In: *Proceedings of the 18th international conference on machine learning (ICML 01)*, pp 577–584
8. Hong Y, Kwong S, Wang H, Ren Q, Chang Y (2008) Probabilistic and graphical model based genetic algorithm driven clustering with instance-level constraints. In: *Proceedings of the IEEE congress on evolutionary computation (CEC 08)*, pp 322–329
9. Hong Y, Kwong S, Xiong H, Ren Q (2008) Genetic-guided semi-supervised clustering algorithm with instance-level constraints. In: *Proceedings of the genetic and evolutionary computation conference (GECCO 08)*, pp 1381–1388
10. Dutta H, Passonneau RJ, Lee A, Radeva A, Xie B, Waltz D, Taranto B (2011) Learning parameters of the K-Means algorithm from subjective human annotation. In: *Proceedings of the 24th international florida artificial intelligence research society conference (FLAIRS 11)*
11. Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65

# Chapter 69

## Traffic Light Detection and Tracking Based on Euclidean Distance Transform and Local Contour Pattern

Zhenyang Wang, Zhidong Deng and Zhen Huang

**Abstract** This paper proposes a new recognition approach for traffic light based on Euclidean distance transform (EDT) and local contour pattern (LCP). There are two main contributions of this paper. First, this paper combines principle component analysis (PCA) with EDT-based image to detect traffic light colors. The color space for specific colors is partitioned more precisely, which leads to a high recognition rate. Second, we incorporate the above color detection into the contour segmentation of traffic light holder based on the LCP to further improve the recognition rate of traffic light. The experimental results show that our approach is able to detect traffic light far away from camera about 50–80 m and the average recognition rate can reach up to 99.29 %.

### 69.1 Introduction

In the urban environment, the traffic light recognition approach is usually used for unmanned ground vehicle (UGV) and advanced driver/pedestrian assistance systems. It has high requirements of real-time, accuracy, and detection distance. For example, if the traffic light recognition approach can not meet the real-time requirements, it will lead to the UGV is not able to make correct response to actual traffic signal situation. If the accuracy is low, the detection failure will probably

---

Z. Wang (✉) · Z. Deng · Z. Huang

State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China  
e-mail: crazycry2010@gmail.com

Z. Deng  
e-mail: michael@tsinghua.edu.cn

Z. Huang  
e-mail: chinaheart2003@qq.com

cause traffic accidents. Additionally, too near detection distance may result in insufficiencies in safety braking distance.

In 2007, Kim proposed a real time traffic light recognition system based on computer vision to improve the safety of color-blind drivers [1]. They developed a simple adaptation method to handle illumination changes over weather and time. But their approach caused a relative high false drop rate. In 2008, the traffic light recognition approach proposed by J. Park et al. from Korea University [2] demonstrated good performance in real time. But they concerned more about traffic light color, and paid insufficient attention to the information on the signal light contour, which indicated low recognition rate. In addition, Lu et al. presented an approach to traffic light detection and classification [3]. Their approach employed the color space to segment the candidate regions of traffic light. Then the template matching was conducted for each candidate region. But the recognition distance of their approach could not be far enough. In 2011, Jan Roters's recognition approach for traffic light took full use of both the color and the contour information [4]. However, their approach was mainly used to detect the crosswalk lights for blind people. In 2011, Chiang et al. proposed a traffic light recognition approach by genetic approximate ellipse detection and spatial texture layouts [5]. Their approach exploited the shape of the traffic light, but they did not take the outer contour of traffic light holder into consideration. In view of the requirements of the UGVs, there is still distinct gap in the performance of traffic light detection. The latter requires that any recognition approach must have not only real-time performance but also high accuracy and far detection distance.

This paper improves the performance of real-time, accuracy, and detection distance. Our approach uses both the color and the shape information to detect traffic lights. First, we combine PCA with EDT-based image to detect colors. The commonly-used approach of color detection is the threshold value determination for each pixel of three color channels. This paper performs a more detailed segmentation in color space, which enhances the detection performance. Second, we incorporate the above color detection into the contour segmentation of traffic light holder based on the LCP to further improve the recognition rate of traffic light. In fact, the traffic light detection based on contour segmentation of traffic light holder is much more reliable compared to the detection based on color alone. The reason is that color is sensitive to the illumination. But the contour segmentation of traffic light holder allows a certain degree of incomplete of the edge image, which makes it more robust.

Our recognition approach, at the vehicle speed of about 35 km/h, can detect traffic light located in front of 50–80 m. The approach can accompany 5–6 frames per second, and the average recognition rate can reach up to 99.29 %, which can meet the required performance of traffic light recognition in real time, accuracy, and detection distance.

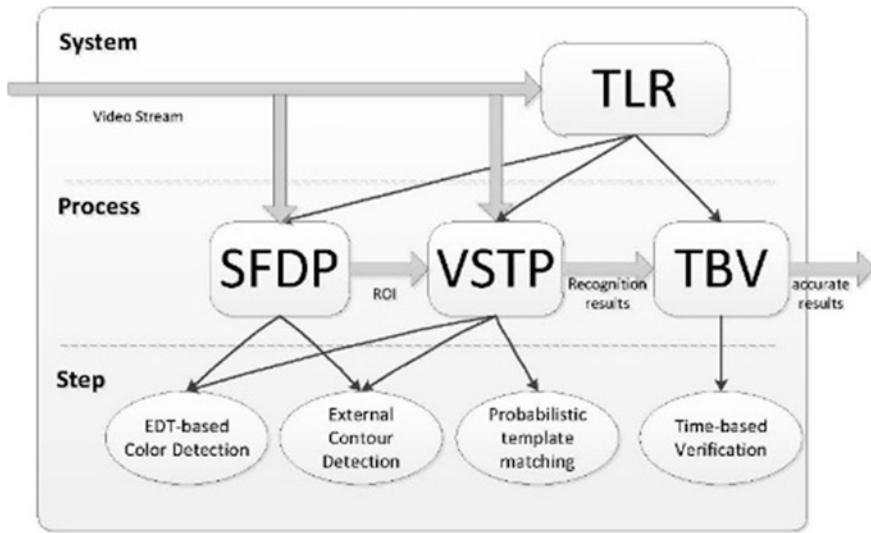


Fig. 69.1 The structure block diagram of traffic light recognition system

## 69.2 A New Traffic Light Recognition Approach

The structure of our traffic light recognition (TLR) system is shown in Fig. 69.1. It consists of two concurrent processes, i.e., the single frame detection process (SFDP) and the video stream tracking process (VSTP), and one time-based verification (TBV). The SFDP only detects the region of interest (ROI) of traffic light candidates, and then deliver them to the VSTP. In the SFDP, it contains both EDT-based color segmentation and contour segmentation of traffic light holder based on the LCP [5]. The VSTP detects the ROIs from the SFDP for the previous frames, then transfer the recognition results to the TBV. The VSTP is similar to the SFDP, but has one more step, i.e., the one based on probabilistic template matching. The TBV helps to improve the reliability of traffic light recognition. It compares the results from the VSTP for at least 10 frames, and draws a relatively accurate result.

### 69.2.1 EDT-Based Color Detection

Color is one of the salient features of the traffic light. The traffic light recognition approach detects all the traffic light candidates through the color segmentation, and then removes all the false detections by contour segmentation of traffic light holder. Thus an accurate color recognition can improve the efficiency of the traffic light recognition.

In fact, the RGB color space is very sensitive to illumination (e.g. due to sunshine, weather, clouds), which is one of the big problems in color segmentation [6]. For this problem, one solution is to use HSV color space. It is believed to be capable of eliminating the impact of illumination intensity.

This paper only focuses on the red and green traffic lights. For each pixel to be detected, in appropriate color space, the Euclidean distance between this pixel and the red (or green) sample pixels is calculated. If the Euclidean distance is less than a certain threshold [7], then the color of this pixel is viewed to be red (or green). When calculating the Euclidean distance, we employ the PCA method in the HSV color space to reduce the dimension, in order to lower the computational complexity.

After reducing the three dimension to the two one, the approach can generate the EDT-based image through the transform matrix of PCA [8]. EDT-based image is the Euclidean distance for each pixel to the nearest target pixel in a 2D image. For each pixel  $p$ , the Euclidean distance of  $p$  is given by

$$t(p) = \min\{d(p, q), q \in S\}$$

where  $t(p)$  denotes the value of the pixel  $p$  in the EDT-based image, which is the closest distance of the pixel  $p$  from the target pixels,  $d(p, q)$  is the Euclidean distance between pixel  $p$  and pixel  $q$ , and  $S$  is the target set of pixels.

In the EDT-based image, the greater the gray of the pixel is, the brighter the illumination intensity is, and the farther the pixel is away from the nearest target pixel. The procedure of establishing the EDT-based image  $m \times n$  is given below.

1. The Euclidean distance between the pixel  $(i, j)$  and the pixel  $(k, t)$  is defined by

$$d = \sqrt{(i - k)^2 + (j - t)^2}$$

2. Establishing a single channel grayscale image of the size  $m \times n$ , and the value of each pixel is 255.
3. For each of the sample, if the sample's location is  $(i, j)$  after mapping to the grayscale image. For each pixel  $(k, t)$  in the grayscale image, calculate the distance  $d$  from the pixel  $(k, t)$  to the sample  $(i, j)$ , it has

$$\text{image}(k, t) = \min\{\text{imgae}(k, t), \sqrt{(i - k)^2 + (j - t)^2}, (i, j) \in S\}$$

For a pixel to be detected, the approach uses the transform matrix of PCA to mapping it from the HSV color space to the EDT-based image. Taking the location of the EDT-based image as  $(i, j)$ , the approach compares the gray value at  $(i, j)$  with the red (or green) threshold. If the gray value at  $(i, j)$  is less than a red (or green) threshold, then the color of this pixel is red (or green).



### 69.2.2 Combination of Color Segmentation and Contour Detection

This paper sequentially uses of both the color and the contour information. The approach detects all the possible traffic light candidates by color segmentation described above. In this place, we aim at removing all the impossible traffic light candidates by contour detection of traffic light holder [9, 10]. After contour detection, the approach gets the ROI that may contain traffic light. Due to the relatively fixed aspect ratio of traffic lights, we could segment each traffic light in the ROI. Consequently, the approach takes use of the probabilistic template matching method to recognise traffic light [11].

First, the approach uses the canny algorithm to extract edge image from the original image. Second, for each pixel of the edge image, we exploit the LCP operator [12] to extract the pixels that belong to the horizontal or the vertical local contour. At last, the approach produces the local contour image that only contains the horizontal and the vertical local contour. The local contour patterns operator used in this paper is shown in Fig. 69.2.

After generating the local contour image, the approach gives rise to the ROI which may cover traffic light. The traffic light holder in the image is in a rectangular shape. Thus it segments the ROI by detecting and locating the rectangular in the local contour image.

The template matching algorithm only concerns the circular and the sagittate samples of traffic light. The size of the template is fixed to  $30 \times 30$  [13]. For each ROI segmented by LCP method, we scale it to the size of  $30 \times 30$  for matching.

To improve the accuracy of the recognition, the approach adopts a state queue to store the last  $n$  ( $n = 10$ ) detection results. Only as the last  $m$  ( $m < n$ ) results are the red (or green) light, the approach gives a recognition of red (or green) traffic signal.

## 69.3 Experimental Results

The experimental platform is shown in Fig. 69.3. This is the new generation of UGV developed by ourselves. The camera is installed on the roof of this UGV called THU-ALV2, which parameters are below: the focal length is 8.5 mm, the resolution is  $1292 \times 964$ , and the frame rate is 32 frame/s.

**Fig. 69.2** The LCP operator [12]

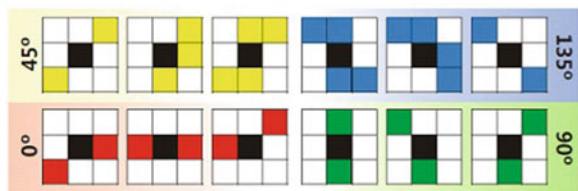




Fig. 69.3 The experimental platform. *Left* is THU-ALV2, *right* is the color camera installed

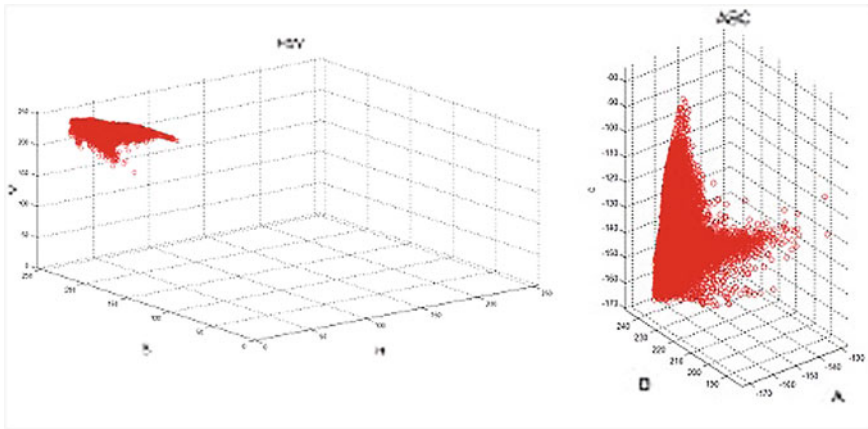
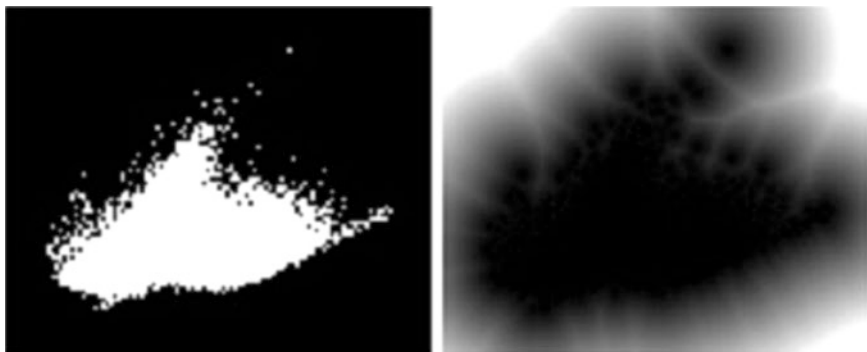


Fig. 69.4 The distribution of the red traffic light training sample set. *Left* is the distribution of the red traffic light training sample set, *right* is the distribution of the red traffic light training sample set after PCA transform

In the color segmentation, the distribution of the red traffic light training sample set is shown in Fig. 69.4. Apparently, the color feature is more significant after PCA transform was done. Meanwhile, the approach reduced the dimension. Then we could achieve the gray image by mapping the PCA transform results to a two-dimensional image. Figure 69.5 gives the gray image and the EDT-based image of the red traffic light training sample set.

After that, we carried out the contour segmentation of traffic light holder. Figure 69.6 gives the results obtained from the LCP transform. The segmentation of the ROI is shown in Fig. 69.7. The probabilistic templates of circular samples and sagittate samples are shown in Fig. 69.8.

In this paper, we conducted a lot of on-site experiments for seven situations on traffic light. The experimental results are shown in Table 69.1. The seven scenarios

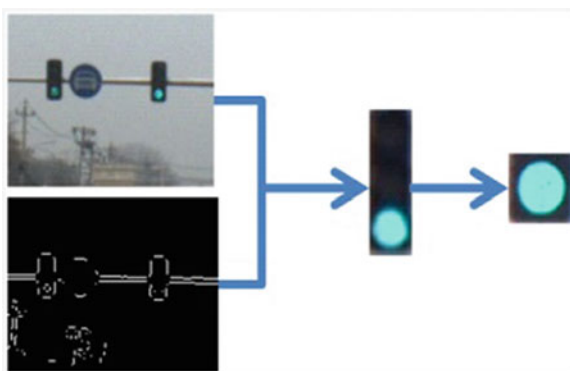


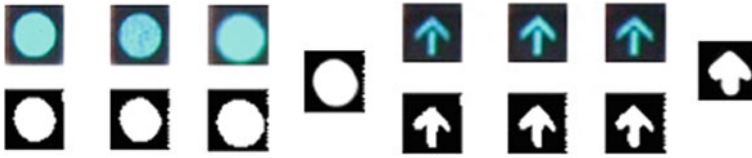
**Fig. 69.5** The gray image of the red traffic light and its EDT-based image. *Left* is the gray image of the red traffic light color training set, *right* is the EDT-based image of the red traffic light color training set



**Fig. 69.6** The results obtained from the LCP operation. *Left* is the original image, *right* is the results obtained from the LCP operation

**Fig. 69.7** The segmentation of the ROI





**Fig. 69.8** The probabilistic templates. *Left* is circular traffic light, *right* is sagittate traffic light

**Table 69.1** The accuracy analysis on different scenarios

	Total frames	False frames	Residual frames	Accuracy (%)	Average accuracy
Scenario #1	1,595	7	0	99.56	99.29 %
Scenario #2	1,442	1	14	98.96	
Scenario #3	340	0	0	100	
Scenario #4	590	1	0	99.83	
Scenario #5	425	0	0	100	
Scenario #6	578	7	0	98.79	
Scenario #7	501	0	9	98.20	

contain four conditions including sunny, cloudy, morning, and afternoon. Scenarios 1, 2, and 3 indicated off-line experiments for traffic light and the remainings on-line ones. In each scenario, we evaluated the total frames of the whole video stream, the false detections counting by frames, the residual errors (undetected errors) counting by frames, and the accuracy. As a result, we calculated the average accuracy over those experiments.

It can be observed from Table 69.1 that the traffic light recognition system has a high detection accuracy. The average recognition rate can reach up to 99.29 %. The proposed approach is able to detect traffic lights far away from camera about 50–80 m. The traffic light recognition system can do real-time recognition. Five or six frames can be processed in one second. The traffic light recognition system has little interference by background such as pedestrians, leading vehicles, and tail-lights of leading vehicles.

## 69.4 Conclusion

This paper improves the performance of traffic light recognition in real-time, accuracy, and detection distance. The proposed approach jointly uses the color and the contour information to detect traffic lights. First, we combine PCA with EDT image to detect colors. Second, this paper incorporates the above color detection into the the LCP to further improve the recognition rate of traffic light. Finally, we conduct a lot of experiments. The results show that our approach is able to detect

traffic lights of far away from camera about 50–80 m and the average recognition rate can reach up to 99.29 %.

In the future, we will further reduce both the time complexity and the space complexity. Some new learning machine approaches such as deep learning and manifold learning should be used for the traffic light recognition problem.

**Acknowledgments** This research work was supported in part by National Natural Science Foundation of China (NSFC) under grant No. 90820305 and No. 60775040.

## References

1. Kim Y, Kim K, Yang X (2007) Real time traffic light recognition system for color vision deficiencies. In: ICMA 2007 International conference on mechatronics and automation, pp 76–81
2. Park JH, Jeong CS (2008) Real-time signal light detection. In: Second international conference on future generation communication and networking symposia FGCNS 08, vol 3, pp 139–142
3. Lu K, Wang C, Chen S (2008) Traffic light recognition. *J Chinese Inst Eng* 31:1069–1075
4. Roters J, Jiang X, Rothaus K (2011) Recognition of traffic lights in live video streams on mobile devices. *IEEE Trans Circ Syst Video Technol* 21(10):1497–1511
5. Chiang C, Ho M, Liao H, Pratama A, Syu W (2011) Detecting and recognizing traffic lights by genetic approximate ellipse detection and spatial texture layouts. *Int J Innovative Comput Inf Control* 769:19–34
6. Fleyeh H (2004) Color detection and segmentation for road and traffic signs. In: IEEE conference on cybernetics and intelligent systems, vol 2, pp 809–814
7. Omachi M, Omachi S (2010) Detection of traffic light using structural information. In: IEEE 10th international conference on signal processing (ICSP), pp 809–812
8. Jiang R, Reinhard K, Tobi V, Wang S (2011) Lane detection and tracking using a new lane model and distance transform. *Mach Vis Appl* 22:721–37
9. Lafuente-Arroyo S, Gil-Jimenez P, Maldonado-Bascon R, Lopez-Ferreras F, Maldonado-Bascon S (2005) Traffic sign shape classification evaluation I: SVM using distance to borders. In: Proceedings of the IEEE intelligent vehicles symposium, pp 557–562
10. Gil-Jimenez P, Lafuente-Arroyo S, Gomez-Moreno H, Lopez-Ferreras F, Maldonado-Bascon S (2005) Traffic sign shape classification evaluation. Part II. FFT applied to the signature of blobs. In: Proceedings of the IEEE intelligent vehicles symposium, pp 607–612
11. Miura J, Kanda T, Shirai Y (2000) An active vision system for real-time traffic sign recognition. In: Proceedings of the IEEE intelligent transportation systems, pp 52–57
12. Landesa-Vazquez I, Parada-Loira F, Alba-Castro J (2010) Fast real-time multiclass traffic sign detection based on novel shape and texture descriptors. In: 13th International IEEE conference on intelligent transportation systems (ITSC), pp 1388–1395
13. de la Escalera A, Moreno L, Salichs M, Armingol J (1997) Road traffic sign detection and classification. *IEEE Trans Ind Electron* 44(6):848–859

# Chapter 70

## Multi-Correlation-Based Mode Decision for Multi-View Video Coding

Qinghong Shen, Fengsui Wang and Sidan Du

**Abstract** In this paper, an efficient mode decision algorithm, named the multi-correlation-based mode decision (MMD), is proposed to reduce the computational complexity while maintaining the high coding efficiency. In this method, the rate distortion (RD) cost of the Direct mode is always computed and compared with an adaptive threshold as a possible early termination chance. This adaptive threshold is determined by using the spatial, temporal and inter-view correlation between the current macroblock (MB) and its neighboring macroblocks. Experimental results demonstrate that the proposed MMD algorithm can significantly achieve computational saving of 72.38 % on average with no significant loss of rate-distortion performance, compared with the full mode decision in the reference software of multi-view video coding (MVC).

**Keywords** Multi-view video coding · Multi-correlation · Full mode decision · JMVC

### 70.1 Introduction

In recent years, with the development in camera and display, the new multimedia applications, such as 3DTV and free-view point TV etc. have been emerging, which can provide people with the highly-welcome experience of 3D stereoscopic

---

Q. Shen · F. Wang (✉) · S. Du (✉)  
School of Electronic Science and Engineering, Nanjing University,  
163 Xianlin Road, Nanjing, China  
e-mail: fengsuiw@163.com

S. Du  
e-mail: coff128@nju.edu.cn

Q. Shen  
e-mail: qhshen@nju.edu.cn

and freedom of selecting the viewpoint [1]. The key of these new multimedia applications is the multi-view video, which is captured by a set of video cameras from various viewpoints but at the same time [2]. In order to efficiently compress and transmit these multi-view videos for practical applications, the Joint Video Team (JVT) of ITU-T VCEG and ISO/IEC MPEG has standardized MVC as a new extension of H.264/AVC (i.e., Annex H) [3]. Joint multi-view video coding (JMVC) has been developed as MVC reference software. In order to achieve higher coding efficiency, JMVC not only adopts intricate intra prediction using spatial correlation and variable block-size motion estimation (ME) using temporal correlation within a single view, but also uses variable block-size disparity estimation (DE) using inter-view correlation between neighbor views, compared with H.264/AVC. However, the higher possible coding efficiency is achieved at the expense of extremely large computation complexity, which obstructs MVC from practical application [4]. Therefore, it is necessary to design an algorithm for reducing computational complexity with maintaining almost the same video coding quality and the total bit rate.

Recently, some fast mode decision method for MVC has been developed to reduce computational complexity in [5–8]. Huo et al. [5] proposed a scalable prediction structure to skip the DE process adaptively for reducing computational complexity in inter-view prediction. A fast mode decision was proposed by Chan et al. [6] for MVC based on a set of dynamic thresholds, which were determined by making use of the on-line statistical analysis of motion and disparity costs of the first group of picture (GOP) in each view. Zeng et al. [7] proposed an early termination scheme for skip mode by checking whether the RD cost of skip mode is below an adaptive threshold for providing a possible early termination chance. According to coding information of the corresponding MB and neighboring MBs in the neighboring view, Shen et al. [8] proposed four fast mode decision algorithms to reduce the computational complexity of ME and DE for MVC.

In this paper, an efficient mode decision algorithm for MVC is proposed. For each current MB, the proposed algorithm always begins with checking whether the RD cost of Direct mode is below an adaptive threshold. If so, the Direct mode will be chosen as the optimal mode and the mode decision process is early terminated. Otherwise, the full mode decision is performed to find the optimal mode. Experimental results have shown that proposed method can greatly reduce the computational complexity while keeping almost the same coding efficiency.

## 70.2 Observation and Motivation

In order to achieve higher coding efficiency, the JMVC employs hierarchical B picture (HBP) prediction structure [9] and a wide set of block sizes that are powerful for exploiting the spatial, temporal and inter-view correlation. For ME and DE, a MB can be partitioned into seven block sizes,  $16 \times 16$ ,  $16 \times 8$ ,  $8 \times 16$ ,  $8 \times 8$ ,  $8 \times 4$ ,  $4 \times 8$ , and  $4 \times 4$ . For the inter-frame MB, there are 11 modes:

**Table 70.1** Distribution of optimal mode in MVC (%)

Sequences	Direct	Seven block sizes	Intra
Flamencol	80.79	18.13	1.08
Race1	83.22	14.58	2.20
Ballroom	77.15	21.37	1.48
Exit	86.84	12.98	0.18
Akko and Kayo	77.83	20.02	2.15
Rena	81.07	15.79	3.14
Uli	65.55	31.50	2.95
Average	78.92	19.20	1.88

Direct,  $16 \times 16$ ,  $16 \times 8$ ,  $8 \times 16$ ,  $8 \times 8$  which includes four sub-modes,  $\text{intra}_4 \times 4$ ,  $\text{intra}_8 \times 8$ , and  $\text{intra}_{16} \times 16$ . In order to maximize the coding efficiency, JMVC adopts the rate distortion optimization (RDO) function [10] and full mode decision to identify the optimal mode. In detail, full mode decision is to exhaustively check all the prediction modes and compute the RD cost of each mode, and then select the one with the minimum RD cost as the optimal mode. Consequently, the computational complexity from the full mode decision is extremely heavy. Thus, a fast mode decision algorithm is highly desirable, especially for practical application of MVC.

Intuitively, the Direct mode is fit for coding those homogeneous regions with motion-less or slow motion. These scenes are often encountered in the nature video sequences. It means that the Direct mode should be more likely to be the optimal mode. To verify this intuition, extensive experiments have been conducted to obtain the distribution of optimal mode by using the full mode decision in JMVC with Quantization Parameter (QP) = 32 based on a set of multi-view video sequences as listed in Table 70.1.

One can see from Table 70.1 that Direct mode is the dominant mode to be the optimal mode for MVC, especially for those homogeneous sequences with slow motion. It should be pointed out that the checking process of the Direct mode occupies very small computational complexity while the following ME, DE and Intra prediction are very time consuming. Therefore, this observation implies that ME, DE and Intra prediction computation of each MB can be entirely saved if Direct mode can be pre-decided. If so, great computational complexity can be reduced.

### 70.3 Proposed Multi-Correlation-Based Mode Decision Method

Based on aforementioned observation, the Direct mode should be pre-decided at the beginning of the mode decision process to provide an early termination opportunity by checking whether RD cost of the Direct mode is below an adaptive threshold. Now the key question is how to determine this adaptive threshold. Since



multi-view video contains a large amount of spatial, temporal and inter-view correlations, the coding information of the current MB are highly correlated to that of its spatial, temporal and inter-view adjacent MBs. As a result, an early Direct mode decision method is developed as follows by making full use of these spatial, temporal and inter-view correlation in multi-view video.

First, it is easily perceived that the current MB (i.e.,  $MB_0$  in Fig. 70.1) has high spatial correlation with its spatial-adjacent MBs (i.e.,  $MB_1$ ,  $MB_2$  and  $MB_3$  in Fig. 70.1). Thus, we can use this spatial correlation to compute the spatial threshold  $T_S$  as:

$$T_S = \sum_{i=1}^3 w_i \cdot B_i \cdot \text{RDcost}(\text{Direct})_i / \sum_{i=1}^3 w_i \cdot B_i \quad (70.1)$$

where  $\text{RDcost}(\text{Direct})_i$  and  $w_i$  are the RD costs of Direct mode and the weights for the corresponding spatial-adjacent  $MB_i$ , for  $i = 1, 2, 3$ . This spatial threshold is designed based on two observations: (1) The closer the adjacent MB to the current MB, the larger the weight should be assigned. Hence, the weights  $w_i$  are empirically determined from the extensive experiments. Here,  $w_1 = w_2 = 1.11$ , and  $w_3 = 0.78$ . (2) The adjacent MB selecting the Direct mode as its optimal mode has

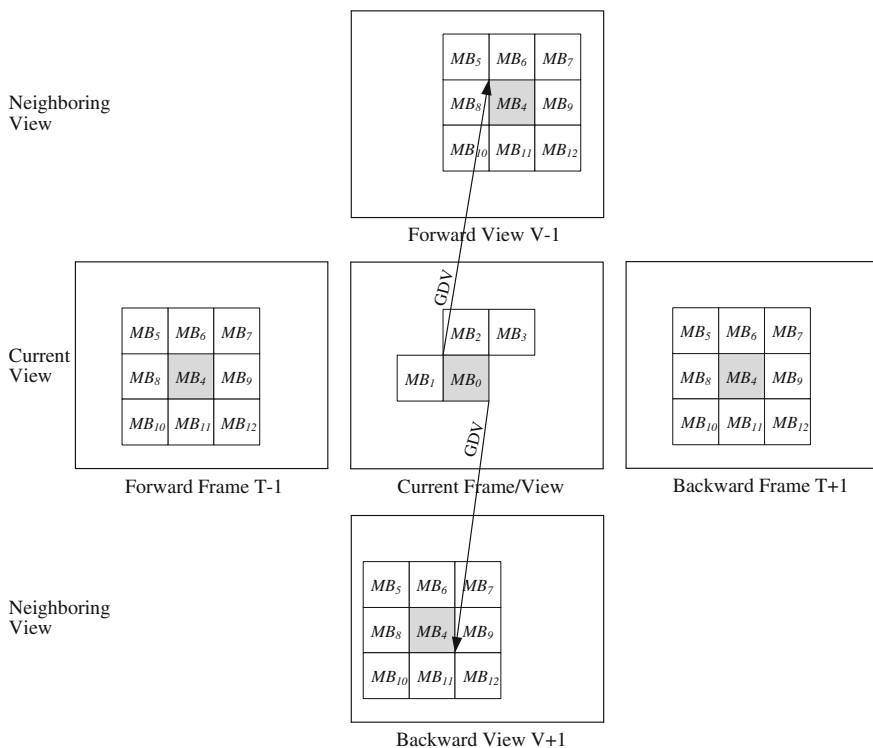


Fig. 70.1 Spatial-adjacent, temporal-adjacent and inter-view-adjacent MBs of the current MB

the reference value to the current MB. In other words, only the RD cost value of the adjacent MB choosing the Direct mode as its optimal mode will be used to compute the threshold. Hence,  $B_i$  is defined as:

$$B_i = \begin{cases} 1 & \text{if the optimal mode of } MB_i \text{ is Direct Mode} \\ 0 & \text{else} \end{cases} \quad (70.2)$$

For the special case that all the  $B_i$  are equal to 0, the spatial threshold  $T_S$  does not exist.

Second, the temporal-adjacent MBs of the current MBs (i.e. Forward Frame  $T - 1$  and Backward Frame  $T + 1$ ) are shown in Fig. 70.1. Note that  $MB_4$  is the MB with the same position as the current MB in forward or backward picture and  $MB_i$   $i = 5, 6, \dots, 12$ , are the 8-neighbors of  $MB_4$ . Considering that the current MB has strong temporal correlation with those temporal-adjacent MBs in both forward and backward frames, the temporal threshold  $T_T$  is computed as the average of the forward temporal threshold  $T_{T-1}$  and the backward temporal threshold  $T_{T+1}$ . Both  $T_{T-1}$  and  $T_{T+1}$  are calculated by fully exploiting the correlation between the current MB and its neighboring MBs. The details of the derivation process of the temporal threshold will be introduced later on, together with that of inter-view threshold.

Third, the inter-view-adjacent MBs of the current MBs (i.e. Forward View  $V - 1$  and Backward View  $V + 1$ ) are shown in Fig. 70.1. It should be pointed out that  $MB_4$  is the corresponding MB of the current MB in forward or backward view, which is identified by using global disparity vector (GDV) [8]. And  $MB_i$   $i = 5, 6, \dots, 12$ , are the 8-neighbors of  $MB_4$ . Since two pictures at the same time instant from two neighboring views usually present the same content, there exists high inter-view correlation between the current MB and its inter-view adjacent MBs. Similar to the temporal threshold  $T_T$ , the inter-view threshold  $T_V$  is also computed as the average of the forward inter-view threshold  $T_{V-1}$  and the backward inter-view threshold  $T_{V+1}$ . Both  $T_{V-1}$  and  $T_{V+1}$  are calculated by following the way of  $T_{T-1}$  and  $T_{T+1}$ .

Here, the temporal threshold  $T_T$  and the inter-view threshold  $T_V$  are computed as:

$$T_T = (T_{T-1} + T_{T+1})/2 \quad (70.3)$$

$$T_V = (T_{V-1} + T_{V+1})/2 \quad (70.4)$$

where

$$T_x = \frac{\sum_{i=4}^{12} w_i^x \cdot B_i^x \cdot \text{RDcost}(\text{Direct})_i^x}{\sum_{i=4}^{12} w_i^x \cdot B_i^x} \quad (70.5)$$

where  $x$  denotes the neighboring picture, such as the forward picture, backward picture etc., i.e.,  $x \in \{T - 1, T + 1, V - 1, V + 1\}$ ,  $\text{RDcost}(\text{Direct})_i^x$  and  $w_i^x$  denote the RD cost of Direct mode and the weights of the  $MB_i$  located in picture  $x$ ,

**Table 70.2** The weights of the temporal-adjacent and inter-view-adjacent MBs

$MB_i$	4	6, 8, 9, 11	5, 7, 10, 12
$w_i^x$	1.48	1.03	0.85

respectively. Moreover,  $w_i^x$  are also empirically determined from the extensive experiments and then documented in Table 70.2.  $B_i^x$  is used to denote whether the  $MB_i$  located in picture  $x$  selects the Direct mode as the optimal mode. If so,  $B_i^x$  is equal to 1, otherwise,  $B_i^x$  is equal to 0. For the special case that all the  $B_i^x$  for picture  $x$  are equal to 0, the corresponding threshold  $T_x$  does not exist.

In the above process, a set of thresholds (i.e.,  $T_S$ ,  $T_T$  and  $T_V$ ) are individually developed by using the spatial, temporal and inter-view correlation. Now the question is how to determine the suitable adaptive threshold  $T$ , which should be dependent on the content of the current MB. Obviously, with larger threshold, more computational complexity can be reduced while more coding efficiency will be degraded, and vice versa. In order to achieve a good trade-off between the computational complexity and the coding efficiency, the adaptive threshold  $T$  is determined as the median prediction value of a set of thresholds, that is:

$$T = \text{median}(T_S, T_T, T_V) \quad (70.6)$$

For the special case that all the thresholds  $T_S$ ,  $T_T$  and  $T_V$  do not exist, the checking of early termination is skipped and full mode decision is performed.

In summary, the proposed MMD algorithm can be depicted as follows:

Step 1: Check whether the current MB is located in an anchor picture. If so, full mode decision is performed on the current MB to identify the corresponding optimal mode; otherwise, go to step 2.

Step 2: Compute the RD cost of the Direct mode (denoted as  $\text{RDcost}(\text{Direct})$ ).

Step 3: Compute a set of thresholds  $T_S$ ,  $T_T$  and  $T_V$ , respectively, and then derive the adaptive threshold  $T$  according to Eq. (70.6).

Step 4: If  $\text{RDcost}(\text{Direct}) < T$ , perform early Direct mode decision, and go to step 6. Otherwise, go to step 5.

Step 5: Perform full mode decision and all the modes are checked to select the one with the minimum RD cost as the optimal mode.

Step 6: The Direct mode is selected as the optimal mode and the mode decision process is early terminated, then go to step 1 and proceed with next MB.

## 70.4 Experimental Results and Discussion

To evaluate the performance, JMVC 8.0 is selected as the experimental platform. The results tested on seven multi-view video sequences are shown in this paper. Note that three views of each test sequence are chosen for experiments. The first

**Table 70.3** Experimental results of two methods: <A> Ref. [7], and <B> the proposed multi-correlation-based mode decision (MMD) algorithm

Sequence	Method	$\Delta$ PSNR (dB)	$\Delta$ B (%)	$\Delta$ T (%)
Flamenco1	<A>	-0.08	-0.59	-64.38
	<B>	-0.06	-0.99	-75.95
Race1	<A>	-0.07	-0.86	-67.91
	<B>	-0.06	-0.81	-78.02
Ballroom	<A>	-0.05	-0.69	-58.23
	<B>	-0.03	-0.98	-67.29
Exit	<A>	-0.10	-1.65	-67.23
	<B>	-0.05	-1.37	-76.58
Akko&Kayo	<A>	-0.06	-0.79	-63.23
	<B>	-0.05	-0.90	-73.09
Rena	<A>	-0.09	-0.74	-59.89
	<B>	-0.04	-1.25	-70.86
Uli	<A>	-0.11	-0.84	-54.25
	<B>	-0.06	-1.12	-64.87
Average	<A>	-0.08	-0.88	-62.16
	<B>	-0.05	-1.06	-72.38

and third views are used as the reference views, respectively. The second view is used as the inter-view views (i.e., the current view). The experimental setup is described as below: (1) Each test sequence is encoded using the HBP prediction structure with a GOP = 12; (2) The QPs are set at 24, 28, 32 and 36, respectively; (3) RDO and CABAC are enabled; (4) The search range of ME and DE is  $\pm 64$ .

In this work, we compare the proposed method with that presented in Ref. [7], and the results are shown in Table 70.3, where “ $\Delta$ PSNR (dB)”, “ $\Delta$ B (%)”, and “ $\Delta$ T (%)” represent PSNR change in dB, bit rate change in percentage and the entire coding time change in percentage, respectively. The positive values represent increments whereas the negative values represent decrements. It should be pointed out that each method is compared with the full mode decision in JMVC.

From the Table 70.3, one can see that the proposed MMD algorithm can significantly reduce the computational complexity while keeping almost the same coding efficiency. It has reduced encoding time about 72.38 % on average. The loss of coding efficiency is negligible for the proposed algorithm: the average PSNR loss is about 0.05 dB, but the bit rate reduction is about 1.06 %. For a better illustration, Fig. 70.2 shows the RD curves of two sequences “Rena” and “Uli” as examples. One can easily see that the proposed method is able to yield almost the same coding efficiency as that of full mode decision in JMVC. Furthermore, compared with the method presented in Ref. [7], 10.22 % higher computational complexity reduction, 0.03 dB PSNR improvement and 0.18 % bit rate reduction is achieved. In summary, the proposed method outperforms Ref. [7] in terms of both coding efficiency maintenance and computation complexity reduction.

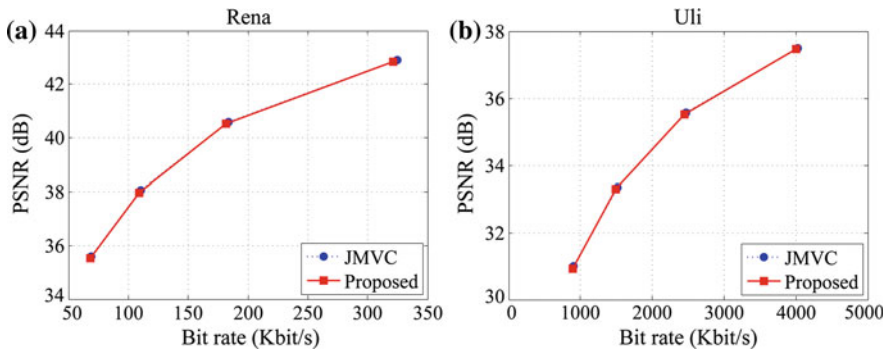


Fig. 70.2 RD curves of two sequences “Rena” and “Uli”

## 70.5 Conclusion

In this paper, an efficient fast mode decision method is proposed for MVC. The computational complexity reduction is achieved by skipping the checking process of the time-consuming ME, DE and intra prediction modes once the RD cost of Direct mode is below an adaptive threshold derived from the multiple correlations in multi-view video. Experimental results have verified that the proposed MMD method is able to significantly reduce the computational load by 72.38 % and the total bit rate by 1.06 % on average, while only incurring a negligible loss of PSNR.

**Acknowledgments** This work was supported by the Major Infrastructure Program of the Health Monitoring System Hardware Platform Based on Sensor Network Node, China (Grant No. 201103051), the National Natural Science Foundation of China under Grand No.61100111, and the Scientific and Technical Supporting Programs of Jiangsu Province, China (BE 2011169).

## References

1. Tanimoto M, Tehrani MP, Fujii T, Yendo T (2011) Free-viewpoint TV. *IEEE Signal Process Mag* 28(1):67–76
2. Muller K, Merkle P, Wiegand T (2011) 3-D video representation using depth maps. *Proc IEEE* 99(4):634–656
3. Wiegand T, Sullivan GJ, Bjonté G, Luthra A (2003) Overview of the H.264/AVC video coding standard. *IEEE Trans Circ Syst Video Technol* 13(7):560–576
4. Vetro A, Wiegand T, Sullivan GJ (2011) Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard. *Proc IEEE* 99(4):626–642
5. Huo JY, Chang YL, Ma YZ (2009) Scalable prediction structure for multiview video coding. In: *IEEE international symposium on circuits and systems*, pp 2593–2596
6. Chan C-C, Lin J-P (2010) On-line statistical analysis based fast mode decision for multi-view video coding. In: *Proceedings of picture coding symposium*, pp 478–481
7. Zeng HQ, Ma K-K, Cai CH (2010) Mode-correlation-based early termination mode decision for multi-view video coding. In: *IEEE international conference on image processing*, pp 3406–3408

8. Shen LQ, Liu Z, An P, Ma R, Zhang ZY (2011) Low-complexity mode decision for MVC. *IEEE Trans Circ Syst Video Technol* 21(6):837–843
9. Merkle P, Smolic A, Muller K (2007) Efficient prediction structure for multiview video coding. *IEEE Trans Circ Syst Video Technol* 17(11):1461–1473
10. Wiegand T, Schwarz H, Joch A, Kossentini F, Sullivan GJ (2003) Rate-constrained coder control and comparison of video coding standards. *IEEE Trans Circ Syst Video Technol* 13(7):688–703

# Chapter 71

## Automated Discrimination of Gait Patterns Based on sEMG Recognition

Fei Wang, Xiao Hao, Baoxiang Zeng, Chucheng Zhou  
and Song Wang

**Abstract** A general scheme of automated discrimination of gait patterns based on recognition of surface electromyogram of lower limbs is proposed to classify three different terrains and six different movement patterns. To verify the effectiveness of different feature extraction methods, time–frequency features such as RMS and MF, wavelet variance and matrix singularity value are employed to process the sEMG signals under different conditions. SVM is used to discriminate gait patterns based on the selected features. Comparison results indicate that feature extraction method based on matrix singularity value can obtain better results and over 92.5 % classification accuracy ratio can be achieved. Experimental result indicates the rationality and effectiveness of the proposed methods for feature extraction and pattern classification. The proposed scheme shows great potential in the application of lower limb assistance.

**Keywords** Gait pattern discrimination · SEMG · Wavelet analysis · SVM

### 71.1 Introduction

Surface electromyogram (sEMG) signal is a comprehensive result of human skin muscle electrical activity in the skin surface time and space. It is the biological electrical signals from human skeletal muscle surface through the non-invasive way to record down the neuromuscular activity by the biological electrical signals,

---

F. Wang (✉) · X. Hao · B. Zeng · C. Zhou · S. Wang  
College of Information Science and Engineering, Northeastern University, 3-11 Wenhua Road, Heping District, Shenyang 110819 Liaoning, China  
e-mail: wangfei@ise.neu.edu.cn

F. Wang  
State Key Laboratory of Robotics and Systems, Harbin Institute of Technology, 92 West Dazhi Street, Nangang District, Harbin 150001 Heilongjiang, China

which can instantaneously reflect neuromuscular function state in the undamaged state. Because sEMG signal control has some direct and natural characteristics, it can be regarded as a source to control intelligent artificial limb. We can use some of the characteristics of multi-channel sEMG for pattern classification, and then drive artificial limb to make different actions. Its application is more and more widely [1]. However, sEMG signal belongs to a kind of stochastic signal and it is a kind of unstable and low signal noise ratio (SNR) of the biological electrical signals, which cannot achieve the ideal effect when using some traditional signal analysis and processing method, so how to obtain the relevant information of body movement from the sEMG signal and complete the human body movement identification of the lower limbs, has become a key problem in myoelectricity prosthesis control field.

## **71.2 Multi-Channel sEMG Acquisition of the Body Lower Limbs**

### ***71.2.1 The Characteristic of sEMG***

sEMG is the neuromuscular system activity of one dimensional time series signal which is recorded down from skeletal muscle surface through the electrode lead. The changes relate to sports unit quantity, motor unit activity pattern and metabolic state and so on. It can be instantly and accurately reflect the muscle activity state and functional status [1, 2]. sEMG is a kind of weak biological signal [3], the amplitude range is in commonly 0–5 mV, muscles contraction 60–300  $\mu\text{V}$  when slack is about 20–30  $\mu\text{V}$ , and generally lower than the level of noise. For health men, myoelectric amplitude of the peak value is 1–3 mV. For body crippled, the muscle electric amplitude is less than 350  $\mu\text{V}$ , some even less than 1  $\mu\text{V}$ , so through the surface electrode the collected sEMG must pass through the amplification and filtering after treatment in order to be more accurate and reliable for scientific analysis and research.

### ***71.2.2 sEMG Signals for Different Gait Patterns***

The human body joint activity is through the central nervous system to control the direction of the physical activity, amplitude, inertia mass and other parameters, so as to produce ideal muscle activity, Due to the research on the table facial electric human lower limb movement classification, muscles in the activities of the lower limbs are mainly rectus femoris, vastus medialis, Vastus lateralis muscle, tibialis anterior, gastrocnemius and so on. These 10 muscles in the lower limb activity contraction are relatively obvious, so we choose the position of the muscle to



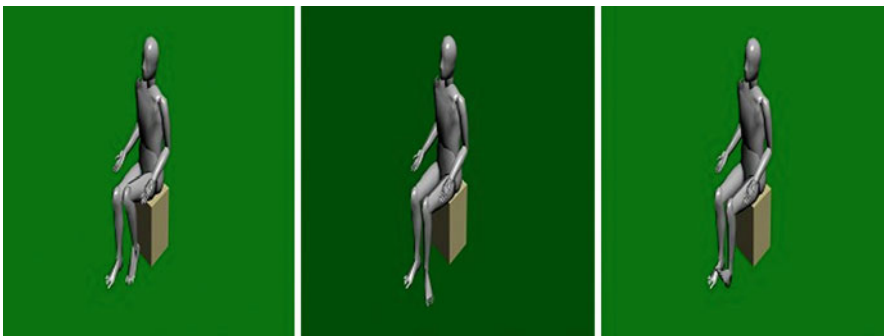
collect the sEMG signal. This paper use the Ag/AgCl biological electrode as lead electrode, every muscle in the direction of the muscle group put two pieces of surface electrode. The diameter of the electrode surface is 10 mm, electrode spacing is about 5–10 cm, we put reference electrode in human body lower limb crural tibial place, because the tibia muscle is in less distribution and can effectively reduce the acquisition process of interference. The electrode position of Vastus lateralis muscle, rectus femoris, vastus medialis is shown in Fig. 71.1.

This paper mainly aims to discriminate common gait patterns in daily life based on recognition of sEMG signals. These gait patterns include level walking, upstairs and downstairs, the movements of squatting kicking and standing up, raising heel, turning one ankle outward and lifting ball flat as shown in Fig. 71.2.

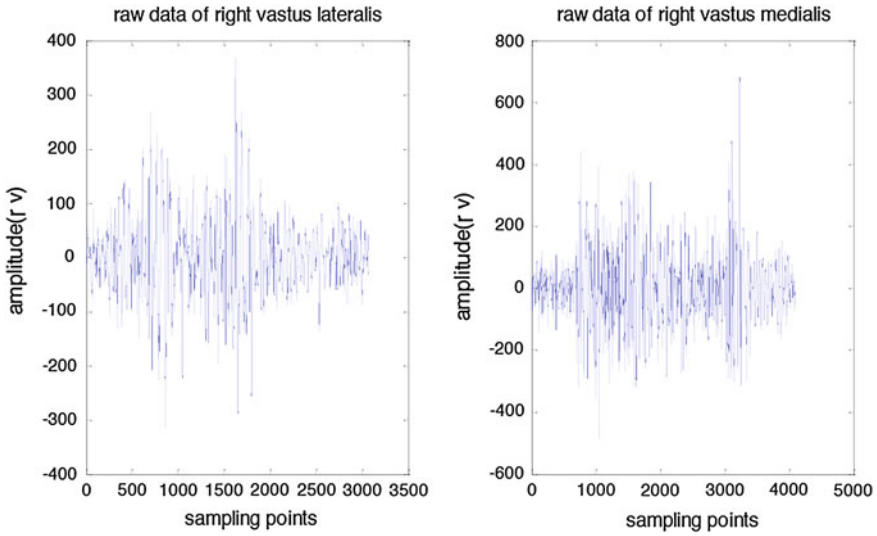
After determining the surface electrode position and movement patterns, through the cable guidance, we put the sEMG signal into the independent development and design of signal conditioning circuit, then magnify and filter the sEMG, so it can remove 50 HZ power frequency interference noise, and keep the 10–500 Hz range of effective sEMG. Using the electromyographic signal acquisition system, we can acquire sEMG signal raw data from the human body right



**Fig. 71.1** Electrode of vastus lateralis, rectus femoris and vastus medialis



**Fig. 71.2** Raising heel, turning one ankle outward and lifting ball flat



**Fig. 71.3** The raw sEMG signals of different lower limb muscles

shares lateral muscle and right shares inside muscle when walking on ground, as shown in Fig. 71.3.

## 71.3 sEMG Signal Feature Extraction

This paper uses time domain, frequency domain and the time–frequency methods to analysis the sEMG signals from lower limbs and wavelet analysis is used to extract the features from the signals.

### 71.3.1 Time Domain Features

In this study, root mean square (RMS) is chosen as the time domain features of sEMG. RMS can reflect sEMG signal amplitude variation characteristics in time dimension. At the same time, it relates to the muscle load resistance factors and muscle itself physiological, the inner relationship between biochemical processes. Therefore, the indicator can be reflect muscle activity state in real-time without damage. The calculation of RMS is as follows:

$$RMS = \sqrt{\frac{\int_t^{t+T} EMG^2(t)dt}{T}} \quad (71.1)$$

RMS curve is directly related to the transformation of muscle force. It is similar to a periodical rhythm curve and is easy to analysis intuitively.

### 71.3.2 Frequency Domain Features

For frequency domain features, spectrum and power spectrum calculated by FFT are selected, which can reflect changes of sEMG signals in different components, therefore, can perfectly reflect changes of sEMG signals in frequency dimension. To depict sEMG signal spectrum or power spectrum quantitatively, the following two indicators, named Mean Power Frequency (MPF) and the Median Frequency (MF) are adopted as shown as follow:

$$MPF = \frac{\int_0^{\infty} fP(f)df}{\int_0^{\infty} P(f)df} \quad (71.2)$$

$$\int_0^{MF} P(f)df = \int_{MF}^{\infty} P(f)df = \frac{1}{2} \int_0^{\infty} P(f)df \quad (71.3)$$

where,  $P(f)$  is muscle power spectrum.

## 71.4 Wavelet Analysis

Wavelet analysis (WA) is a classical local time–frequency analysis method. In WA, the size of windows is fixed, whereas time window and frequency windows are changeable. For example, it has higher time resolution and lower frequency resolution in high frequency part, vice versa. This characteristic makes WA self-adaptable to signals. In small scale, the high frequency character can be expressed, while in large scale, low frequency character can be demonstrated. In this paper, different scales are selected to descript sEMG signals.

The discrete wavelet transform can be expressed as follows:

$$W_f(m, n) = a_0^{-\frac{m}{2}} \int_{-\infty}^{\infty} f(t) \bar{\psi}(a_0^{-m}t - nb_0)dt \quad (71.4)$$

sEMG signal is a kind of unstable stochastic signal, furthermore, it has sensitive dependence on muscle's activity. Fourier transformation is difficult to reflect the change rule of sEMG signal and essence accurately, but wavelet transformation is similar to a group of bandpass filters whose bandwidth is equal and center frequency is variable. In the time domain and frequency domain it has good localization properties and can be used for analyzing the unstable sEMG, considering the small wave of orthogonality and tight branch, Db wavelet is better than the

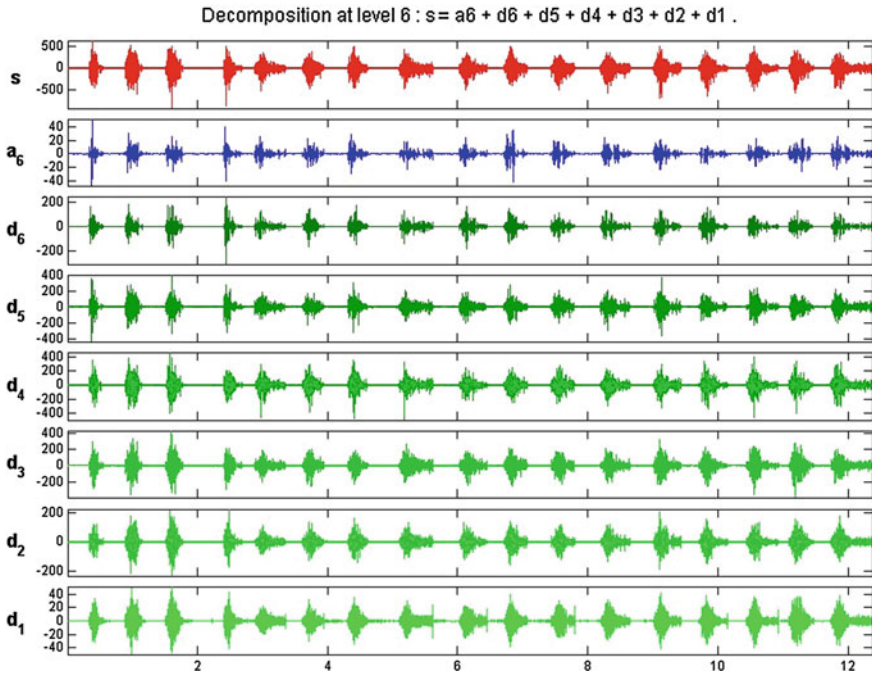


Fig. 71.4 Wavelet decomposition at six level of sEMG

other series of wavelet and at present using Db4 wavelet to analyse sEMG is relative mature [4, 5] decomposition scale generally choose between 5 and 8. This paper choose db4 wavelet to make sEMG signal six scale wavelet decomposition, the wavelet decomposition result is shown in Fig. 71.4.

Aiming at the classification of human lower limb movement mode [6–11] in this paper, we can get all kinds of sports mode sEMG signal wavelet decomposition coefficient matrix singular value and make it as the feature vector for pattern classification.

## 71.5 Gait Patterns Discrimination

Support Vector Machine (SVM) is a kind of statistics theory on the basis of the development of novel learning algorithm. Based on the structural risk minimization principle and in the light of limited samples we are able to find out the global optimal solution, largely solved the model selection, nonlinear, local minimum point and dimension disaster problems.

### 71.5.1 Terrains Discrimination

According to sEMG signal which was acquired from lower limb muscles that human body is in different terrains, we can discriminate three terrains: level walking, upstairs and downstairs. Here, we select sEMG signal of the lateral thigh muscle for analysis. Based on SVM, the three kinds of terrain are classified. The result of classification between level walking and upstairs is shown in Fig. 71.5.

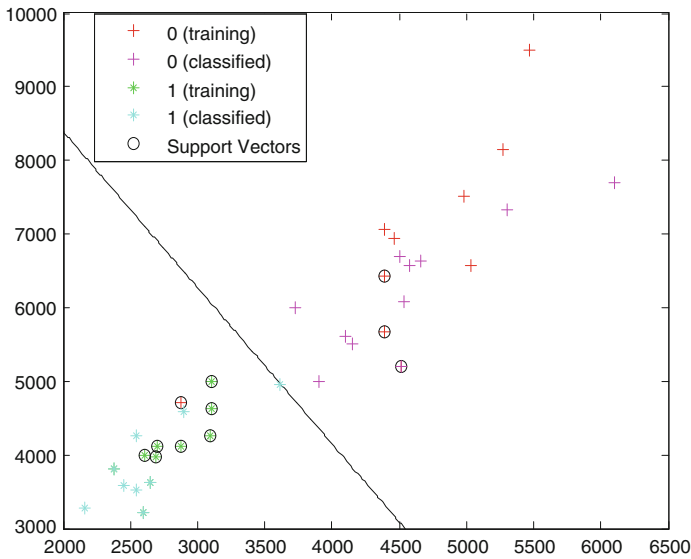
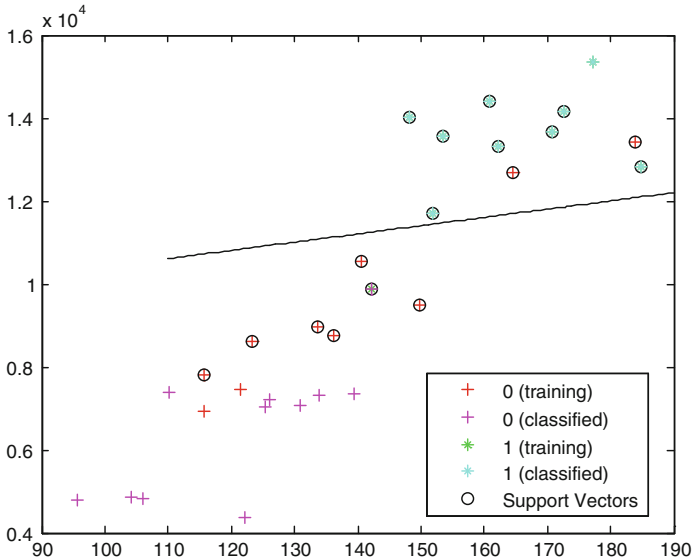


Fig. 71.5 Classification result between level waking and upstairs

### 71.5.2 Movements Discrimination

In this paper, by analysing sEMG signal from lower limb muscles, we completed six movement patterns classification of squatting, kicking and standing up under standing condition, raising heel, turning one ankle outward and lifting ball flat, under sitting situation. By experimental certification, the effect of thigh muscle in the last three action is small, so we select the lateral thigh muscle’s sEMG signal for analysis when making the first three actions in standing condition and we choose the tibial anterior muscle’s sEMG signal for analysis when making the last three actions under sitting situation. Based on the SVM method, two different movements make classification respectively. Classification results of kicking forward and squatting is shown in Fig. 71.6.



**Fig. 71.6** Classification result between kicking forward and squatting

## 71.6 Conclusions

In this paper, two analysis methods were used to extract features from sEMG signals under conditions of three different terrains, and six movement patterns of lower limbs. These two methods are: (1) time-frequency analysis, which use RMS, MF to consist of two-dimensional feature vector; (2) wavelet analysis, which process the acquired sEMG signals by six scale wavelet decomposition, then feature vectors are composed of variance of the wavelet coefficient matrix.

In this paper, it is proposed that matrix singular value extracted from wavelet transform coefficient matrix as signal features, constructing six dimensional feature vectors by the wavelet coefficient matrix singular value, and inputting to the classifier to realize gait patterns classification.

Wavelet transform coefficient matrix singular value has two important features:

- (1) Matrix singular value has a good stability, when the matrix elements changed, the change of singular value itself will be very small;
- (2) Singular value fully reflects the matrix contains information, which is the intrinsic feature of matrix.

Singular value decomposition has been widely used in signal processing, such as confirming the order number in system identification and parameter extraction.

Experimental results show that the feature extraction method based on matrix singular value has higher classification accuracy with the same classifier SVM. The result of classification accuracy of all kinds of feature extraction methods are shown in Table 71.1.

**Table 71.1** Classification results for different feature extraction methods

Feature extraction	Terrains (%)	Standing phase (%)	Sitting phase (%)
MF RMS	87	82	85
Wavelet variance	90	85	95
Singular value	95	90	95

From the table, we can draw the conclusion that the proposed sEMG signal feature extraction method based on matrix singular value is more efficient than traditional time–frequency method and can obtain a more ideal effect.

**Acknowledgments** This work is partially supported by State Key Laboratory of Robotics and System (HIT) Grant #SKLRS-2010-ZD-03 and Research Funds for the Central Universities Grant #N110804005, #N120204002.

## References

1. Massimo S (2009) A lower limb EMG-driven biomechanical model for applications in rehabilitation robotics. In: Proceedings of the 14th International Conference on Advanced Robotics 2:905–911
2. Balestra G, Frassinelli S (2001) Time-frequency analysis of surface myoelectric signals during athletic movement. *IEEE Trans Med Biol* 20(6):106–115
3. Xie H, Wang Z (2006) Mean frequency desired via Hilbert-Huang transform with application to fatigue EMG signal analysis. *Comput Methods Programs Biomed* 82:114–120
4. Zhou P, Lowery MM (2007) Decoding a new neural-machine interface for control of artificial limbs. *J Neurophysiol* 98(5):2974–2982
5. Chen ZW, Hu TP (2002) A reconstruct digit by transplantation of a second toe for control of an electromechanical prosthetic hand. *Microsurgery* 22:5–10
6. Milica MJ (2010) An EMG system for studying motor control strategies and fatigue. In: Proceedings of the 10th Symposium on Neural Network Applications in Electrical Engineering, pp 15–18
7. Macro AC, Garcia C (2010) An alternative approach in muscle fatigue evaluation from the surface EMG signal. In: 32nd Annual international conference of the IEEE engineering in medicine and biology society, pp 2419–2422
8. Balestra F (2001) Time-frequency analysis of surface myoelectric signals during athletic movement. *IEEE Trans Med Biol* 20(6):106–115
9. Ai-Assaf Y, Ai-Nashash H (2001) Myoelectric signal segmentation and classification using wavelets based neural networks. In: Proceedings of the 23rd annual international conference (EMBS), pp 1820–1823
10. Nazarpour K, Sharafat AR, Firoozabadi SMP (2005) Surface EMG signal classification using a selective mix of higher order statistics. In: Proceedings of the 27th annual conference of the engineering in medicine and biology society (IEEE 2005), pp 4208–4211
11. Haight JM (2005) The sensitivity of autoregressive model coefficient in quantification of trunk muscle fatigue during a sustained isometric contraction. *Int J Ind Ergon* 35(4):321–330

# Chapter 72

## Hierarchical Sparse Representation for Traffic Sign Recognition

Yaxiang Fan, Hao Sun, Shilin Zhou and Huanxin Zou

**Abstract** Researchers have proposed various machine learning algorithms for traffic sign recognition (TSR), which is a supervised multiclassification problem with unbalanced class frequencies and various appearances. This paper presents a novel framework for traffic sign recognition exploiting the sparse property of intrinsic information of traffic sign. The contributions of our work are twofold: on one hand, the intrinsic discriminating information among different categories is utilized, on the other an efficient hierarchical sparse representation classification (HSRC) strategy is adopted. Experiments on publicly available datasets show that HSRC is efficient for traffic sign recognition, achieving higher accuracy than many state-of-the-art schemes.

**Keywords** Traffic sign recognition · Sparse representation · Hierarchical classification

### 72.1 Introduction

Traffic sign recognition is an important part in driver assistance system, which aims to display helpful information to the driver using knowledge about the current conditions on the road. Traffic signs represent the current traffic situation, show danger, give warnings to drivers and so on. Characterized by wide variability in their visual appearance in real-world environments, and traffic sign recognition is a multiclassification problem with unbalanced class frequencies. Illumination changes, partial occlusions, rotations, and weather conditions further increase the range of variations in visual appearance a classifier has to cope with.

---

Y. Fan · H. Sun (✉) · S. Zhou · H. Zou  
School of Electronic Science and Engineering, National University of Defense Technology,  
Changsha 410072, Hunan, People's Republic of China  
e-mail: clhaosun@gmail.com



To solve the recognition problem, various machine learning algorithms have been proposed, including template matching [1], neural networks [2], support vector machine [3], etc.

Sparse representation is proposed based on compressed sensing [4]. It aims to find a new sparse space using different bases from the original space. In the sparse space, an image could be represented with a significantly smaller number of bases than when it is represented in original space. The sparsity can help work well in tasks such as image reconstruction, classification, visualization, etc. Recently a work on face recognition [5] showed that sparse representation is naturally discriminative and robust to noise, occlusion.

For traffic signs, which have similar shapes, colors, patterns and texts, are supposed to share certain inherent structures and have inherent sparsity. Furthermore, traffic signs with the same general meaning, such as various speed limits, have a common general appearance, leading to subsets of traffic signs that are very similar to each other.

Motivated by the above observations, we proposed a novel hierarchical sparse representation classification (HSRC) framework for traffic sign recognition. Firstly, an input testing sample is then represented as the linear combination of all the training samples by  $l_1$ -regularized least square method. Then, recognition is accomplished by using a discriminating function defined on the representation coefficients. As the traffic signs have the subsets information, we design a hierarchical classification algorithm to recognize them.

The rest of the paper is organized as follows: [Sect. 72.2](#) describes the methodology. The SR of traffic signs samples is firstly presented, and the algorithm of HSRC is then given in details. [Section 72.3](#) shows the experimental results. [Section 72.4](#) concludes the paper and outlines several directions for future work.

## 72.2 Methodology

In this section, we present the basic components of our hierarchical sparse representation algorithm for traffic sign recognition.

### 72.2.1 Sparse Representation of Traffic Sign

Sparse representation aims to find a new sparse space using different bases from the original space. In the sparse space, an image could be represented with a significantly smaller number of bases than when it is represented in original space. For traffic signs, which have similar shapes, colors, patterns and texts, are supposed to share certain inherent structures. For example, various speed limits have a common general appearance, leading to subsets of traffic signs that are very similar to each other. So we suppose that traffic signs have inherent sparsity.

We arrange the  $n_i$  samples of the  $i$ th class as a matrix  $A_i = [c_{i,1}, c_{i,2}, \dots, c_{i,n_i}] \in \mathbb{R}^{m \times n_i}$  with each sample being a column. Given that the training samples of the  $i$ th class are sufficient, any new (testing) sample  $y \in \mathbb{R}^m$  in the same class will approximately lie in the linear span of the training samples associated with class  $i$ :

$$y = \alpha_{i,1}c_{i,1} + \alpha_{i,2}c_{i,2} + \dots + \alpha_{i,n_i}c_{i,n_i} \quad (72.1)$$

for some scalars  $\alpha_{i,j} \in \mathbb{R}^m$ ,  $j = 1, 2, \dots, n_i$ .

For tumor classification, the membership  $i$  of the new testing sample  $y$  is unknown. We arrange the training data samples of each class in matrix  $A$ . Suppose that the samples with the same class are conjoint, i.e.,  $A = [A_1, A_2, \dots, A_k]$ , then the linear representation of  $y$  can be rewritten in terms of all the training samples as

$$y = Ax_0, \quad (72.2)$$

where, ideally,  $x_0 = [0, \dots, 0, \alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,n_i}, 0, \dots, 0]^T \in \mathbb{R}^n$  is a coefficient vector whose entries are zero except for those associated with the  $i$ th class. In other words, the nonzero entries in the estimate  $x_0$  will be associated with the columns of  $A$  from a single object class  $i$  so that we can assign the testing sample  $y$  to that class.

Now, the key problem to be solved is how to calculate  $x_0$ . It can be expressed as the following optimization problem:

$$\hat{x}_0 = \arg \min \|x\|_0 \text{ subject to } Ax = y \quad (72.3)$$

Finding the solution to the above SR problem is NP-hard due to its nature of combinatorial optimization. Fortunately, recent development in the theory of SR and compressive sensing reveals that if the solution being sought is sparse enough, the solution to the  $l_0$ -minimization problem in (72.3) is equivalent to the solution to the following  $l_1$ -minimization problem:

$$\hat{x}_1 = \arg \min \|x\|_1 \text{ subject to } Ax = y \quad (72.4)$$

This problem can be solved in polynomial time by standard linear programming methods [6].

### 72.2.2 Hierarchical Sparse Representation for Traffic Sign Recognition

Traffic signs can be classified into some subsets with specific similar attributions, such as speed limit signs, mandatory signs, etc. The intra-class discriminative information is quite small, while the between-class discriminative information is big. Such information can be used as the prior knowledge, which is utilized to further recognition.

It has been shown that adding structural constraints to recognition has value both at the level of recognition speed and at the level of recognition accuracy. One intuitive idea is to adopt a divide-and-conquer strategy, where the training samples are partitioned into several subcategorization via a high-level knowledge. To utilize both intra-class intrinsic structure and between-class discriminative information, we propose a hierarchical classification algorithm, called hierarchical sparse representation classification (HSRC). Firstly, we classify the query sign to the subsets with SRC, then we further use SRC to improve recognition accuracy. The procedures of HSRC are summarized in Table 72.1. The structure of HSRC is shown in Fig. 72.1.

### 72.3 Experimental Results

To evaluate our method, we use German traffic sign recognition benchmark (GTSRB) dataset [7] and Belgium Traffic Sign Classification Benchmark (Belgium TSC) dataset [8].

The GTSRB Challenge [7], which was held by IJCNN 2011, is a multi-class, single-image classification challenge. GTSRB has 43 classes with 39,209 training images and 12,630 test images.

Belgium TSC is a subset of the Belgium Traffic Sign Benchmark, which is built for traffic sign classification purposes, and contains 62 different classes of traffic signs. Belgium TSC consists of a training set with 4,591 images and a testing set with 2,534 images.

The proposed HSRC is compared with four state-of-the-art methods: SVM, LRC, SRC, and a typical neural network method, i.e., CCNNs. We use the results of CCNNs in [7] to compare with other methods tested on the same data set. In order to improve the algorithm speed, we extract the metasample from the training samples as the dictionary to represent the test sample. A metasample is a linear

**Table 72.1** The HSRC algorithm

---

Suppose that we have  $k$  subsets, and let  $A = [A_1, A_2, \dots, A_K]$ . Each subsets have  $K_c$  classes of subjects, and let  $A_c = [A_1^c, A_2^c, \dots, A_{K_c}^c]$ ,  $i = 1, \dots, K$ , then

1. Normalize the columns of  $A$  have unit  $l_2$ -norm, code  $y$  over  $A$  via  $l_1$ - minimization  
 $\hat{x} = \arg \min \|x\|_1$  subject to  $Ax = y$  (72.5)

2. Compute the residuals

$$e_j(y) = \|y - A_j \hat{x}_j\|_2 \quad (72.6)$$

where  $\hat{x}_j$  is the coding coefficient vector associated with class  $j$

3. Get the  $c$  subset of  $y$  as

$$c = \text{identity subset}(y) = \arg \min_c \{e_c\} \quad (72.7)$$

4. Then code  $y$  over  $A_c$  via  $l_1$ - minimization

$$\hat{\alpha} = \arg \min \|\alpha\|_1 \text{ subject to } A_c \alpha = y \quad (72.8)$$

5. Output:

$$\text{identity}(y) = \arg \min_i \|y - A_c \hat{\alpha}\| \quad (72.9)$$


---

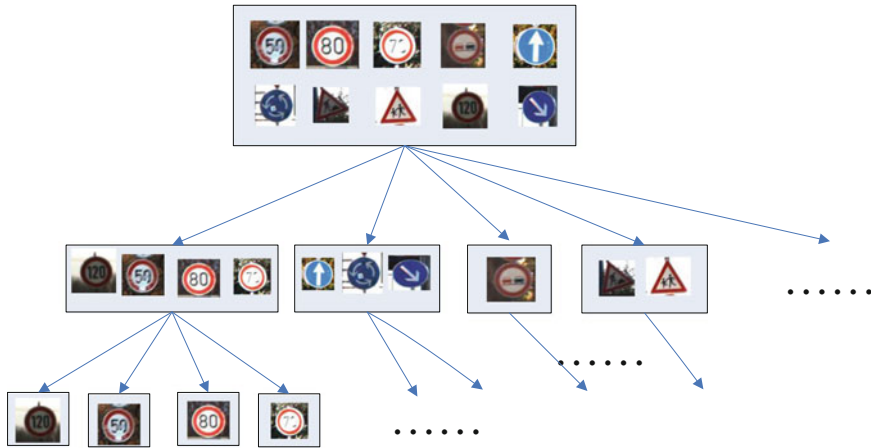


Fig. 72.1 Structure of HSRC

combination of the gene expression profiles of samples, which can capture the alternative structures inherent to the data [9].

### 72.3.1 The GTSRB Dataset

In this experiment, we use HOG Set 2 as the basic feature. The traffic signs of GSTRB can be separated into six large subsets (speed limit, other prohibitory, derestriction, mandatory, danger, and unique signs), which are shown in Fig. 72.2.

All the 43 classes of GTSRB are used to test these methods. The recognition accuracy results are shown in Table 72.2.

From Table 72.2, the experimental results show that HSRC performs better than the other methods except CCNNs. Comparing with SRC, our HSRC

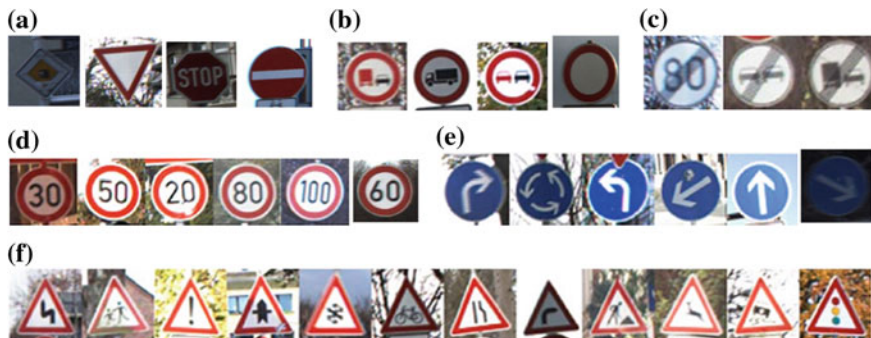


Fig. 72.2 Samples of the six subsets of similar traffic signs of GTSRB. a Unique. b Other prohibitory. c Derestriction. d Speed limit. e Mandatory. f Danger

**Table 72.2** Recognition accuracy results (%) of GTSRB

Methods	SVM	SRC	LRC	CCNNs	HSRC
Accuracy rate	85.42	89.27	90.74	99.42	97.36

contribute to recognition accuracy and recognition efficiency. Although CCNNs exhibit superior performance over classification methods, a high computing cost is paid for neural-network methods to train the data set.

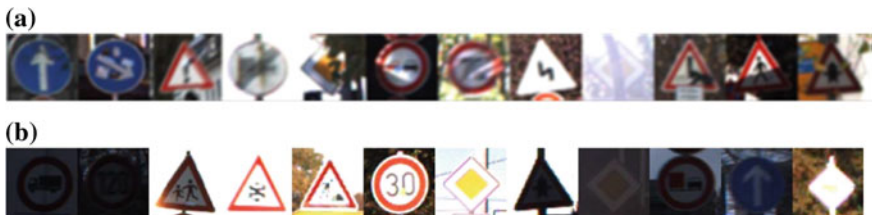
We select examples that are difficult to recognize from the entire test data set to form two new test subsets, corresponding to partial occlusion, various lighting. Some samples are shown in Fig. 72.3. We compare HSRC with three classification methods: SVM, META + SVM, SRC. The individual results for each new subset are listed in Table 72.3.

HSRC achieves considerably better performance than the other methods in three situations. Sparse representation is an effective technique when the signs are under partial occlusion or various lighting conditions. Thus, SRC and HSRC perform better than SVM and LRC. Given that HSRC is also an algorithm that improve sparse representation, it can more effectively address various appearances.

### 72.3.2 The Belgium TSC Dataset

In this experiment, the gray-value feature from the BelgiumTSC is used. The gray-value feature is extracted from the original traffic sign image and resized into  $40 \times 40$  so that it can be represented as a 1600-D vector. The traffic signs of the Belgium TSC can be separated into five large subsets (see in Fig 72.4).

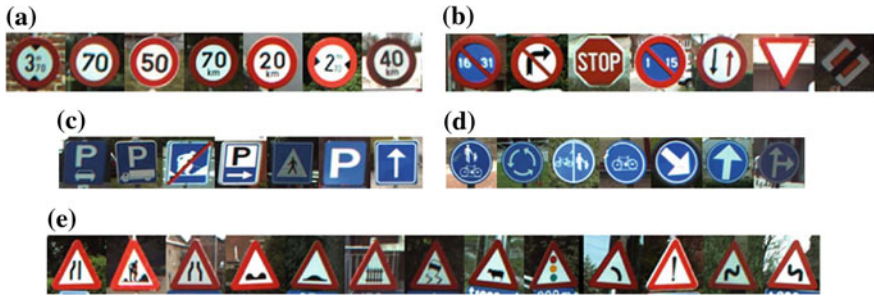
We use the same setup for the experiment as in Sect. 3.1 and the recognition accuracy results are shown in Table 72.4. It can be seen that our HMSRC outperforms the other three methods.



**Fig. 72.3** Samples of the new subsets of GTSRB with various appearances. **a** Traffic signs with partial occlusion. **b** Traffic signs with various lighting conditions

**Table 72.3** Recognition accuracy results (%) of new subsets of GTSRB with various appearances

Methods	SVM	SRC	LRC	HSRC
Partial occlusion	62.31	73.43	67.54	80.61
Various lighting	58.71	70.47	69.33	77.64



**Fig. 72.4** Samples of the six subsets of similar traffic signs of GTSRB. **a** Limit signs. **b** Unique signs. **c** Circle mandatory signs. **d** Rectangle mandatory signs. **e** Danger signs

**Table 72.4** Recognition accuracy results (%) of Belgium TSC

Methods	SVM	SRC	LRC	HSRC
Accuracy rate	80.37	86.83	88.74	92.56

## 72.4 Conclusion

In this study, we proposed a new hierarchical SR-based approach for traffic sign recognition. Classification is achieved by a discriminating function of the SR coefficients, which are obtained by  $l_1$ -regularized least square optimization. To speed up the recognition process, a hierarchical structure is adopted. The proposed hierarchical SR classification (HSRC) was compared with several state-of-the-art machine learning algorithms on the GTSRB dataset and the BelgiumTSC dataset. Experimental results demonstrate the robustness and effectiveness of our proposed method.

**Acknowledgments** This work was supported by Specialized Research Fund for the Doctoral Program of Higher Education (No. 20124307120013).

## References

1. Gu Y, Yendo T, Tehrani M, Fujii T, Tanimoto M (2010) A new vision system for traffic sign recognition. *IEEE Intell Veh Symp* 7–12
2. Ohgushi K, Hamada N (2009) Traffic sign recognition by bags of features. In: *Proceedings of the IEEE Tencon*, pp 1–6
3. Maldonado-Bascón S, Lafuente-Arroyo S, Gil-Jimenez P, Gomez-Moreno H, Lopez-Ferreras F (2007) Road-sign detection and recognition based on support vector machines. *IEEE Trans Intell Transp Syst* 8(2):264–278
4. Donoho D-L (2006) Compressed sensing. *IEEE Trans Inf Theor* 52(4):1289–1306
5. Wright J, Yang A-Y, Ganesh A, Sastry S-S, Ma Y (2009) Robust face recognition via sparse representation. *IEEE Trans Pattern Mach Intell* 31(2):210–227
6. Chen S-S, Donoho D-L, Saunders M-A (2001) Atomic decomposition by basis pursuit. *SIAM Rev* 43(1):129–159
7. Stallkamp J, Schlipsing M, Salmen J, Igel C (2012) Man versus computer: benchmarking machine learning algorithms for traffic sign recognition. *Neural Netw* 32(8):323–332
8. Timofte R, Zimmermann K, Gool L-V (2009) Multi-view traffic sign detection, recognition, and 3rd localisation. In *IEEE workshop on applications of computer vision (WACV)*
9. Xu M, Zheng C, Zhang L, Vincent N (2009) Tumor classification via sparse representation based on metasample. In: *IEEE international symposium on knowledge acquisition and modeling (ISKAM)*

# Chapter 73

## A New Sequence-Based Approach for XML Data Query

Wen Li, Jin Yang, Gaofeng Sun and Sen Yue

**Abstract** In order to avoid expensive join operations in query processing from structured XML document, some index methods based on sequence have been proposed, which transform XML documents and twig patterns into sequences. By performing subsequence matching, query is processed holistically without breaking the twig pattern into many individual root-to-leaf paths, and large useless intermediate results and expensive join operations are avoided. In this paper, combining path sequence strategy with region labeling scheme, we propose a new sequence scheme, Region Path sequence scheme, where the last node of each path is labeled with the region labeling scheme. Compared with previous approaches, our approach can avoid false alarm more effectively, and any extra structure for labeling needn't be constructed. Furthermore, we construct two level B<sup>+</sup>-tree structure to finish the matching, and also propose corresponding matching algorithm. Experiment results demonstrate that our approach can not only avoid false alarm, but also process query more quickly than previous methods such as ViST, and Constraint Sequence.

**Keywords** XML · Sequence structure · False alarm · Region Path

---

W. Li · J. Yang (✉) · G. Sun  
College of Mathematics, Taiyuan University of Technology, Taiyuan 030024, China  
e-mail: ya\_jin100@163.com

S. Yue  
College of Electronic and Information Engineering, Civil Aviation University of China,  
Tianjin 300300, China



### 73.1 Introduction

As a new standard language for representing and exchanging data on the internet, XML [1] is applied in many commercial and scientific fields. As an instance of semi-structured data, XML has very flexible structure, which is often represented as a labeled directed graph. Figure 73.1 shows an example of graph structure.

In order to holistically process XML queries, several sequence-based XML query approaches [2–6] have been proposed, where query expressions are denoted as twig patterns which are tree structures and precisely describe the relationships between two nodes. For instance, the XQuery [7] expression “/Book [// Author = ‘Jane’]” can be presented as a twig pattern in Fig. 73.2. Here, twig pattern are considered to be a basic query unit, which avoids breaking the twig pattern into many individual root-to-leaf paths. In sequence-based query schemes, XML document and XML queries are transformed into sequence structures, and XML queries are processed by subsequence matching.

Wang et al. proposed an approach named ViST, which transforms XML data and XML queries into structure-encoded sequence by preorder traversal of an XML document tree [8]. The structure-encoded sequence is a two-dimensional sequence of (Symbol, Prefix) pairs, where Symbol is a node in the XML document tree and Prefix is the path from the root node to Symbol. The *S1* in Fig. 73.3 is described as  $S1 = ((P, \epsilon) (Q, P) (T, PQ) (S, PQ) (R, P) (U, PR) (T, PR))$ .

ViST gives us a new view of holistic processing of a twig pattern without breaking the twig pattern into root-to-leaf paths and processing these paths individually. However, query processing may result in false alarm. For instance, the *S2* and *Q* in Fig. 73.3 can be presented with ViST sequences  $((P, \epsilon) (Q, P) (T, PQ) (Q, P) (S, PQ))$  and  $((P, \epsilon) (Q, P) (T, PQ) (S, PQ))$  respectively. Although sequence of query *Q* is a subsequence of *S1* and *S2*, the twig pattern *Q* only occurs in *S1* and the matching for *S2* will result in false alarm. False alarm is processed by expensive join operation [8].

PRIX is a more succinct encoding sequence [3]. It constructed a one-to-one correspondence between a labeled tree and a sequence by repeatedly deleting the leaf node with the smallest label and adding the label of its parent to the sequence. With tree nodes labeled by unique postorder numbers and unique ordering of the

Fig. 73.1 A structural XML document

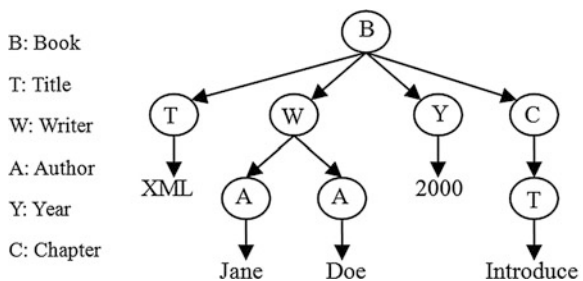
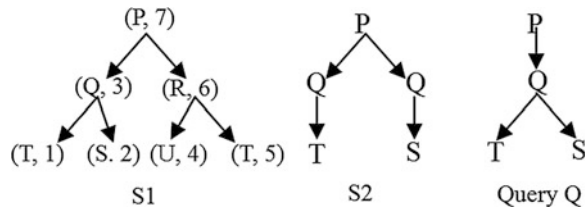


Fig. 73.2 Query twig pattern



Fig. 73.3 An example of matching



numbers, the PRIX sequence for the *S1* of Fig. 73.3 is  $(Q Q P R R P)$ . With unique ordering of the PRIX sequence, PRIX overcomes the false alarm problem.

Instead of using a two-dimensional component in ViST, Constraint Sequence [5] proposed by Wang et al. encodes each node *n* in the tree by the path from the root node to *n*. So, the sequence of *S1* in Fig. 73.3 can be presented as  $\{P, PQ, PQT, PQS, PR, PRU, PRT\}$ .

In constraint sequence, the notion of constraint is introduced, which checks the relationship between any two nodes. Constraint matching is used to avoid false alarm caused by identical sibling nodes and to maintain query equivalence. However, the process of checking identical sibling-cover has a bad effect on query performance.

In this paper, motivated by false alarm and poor query performance, we propose a new sequence scheme, Region Path (RP) sequence scheme which is built by combining sequence based on paths [6] with region labeling scheme [9]. Region Path sequence scheme avoids false alarm and gets precise matching results.

The rest of the paper is organized as follow. In Sect. 73.2, we present our idea. We describe our algorithm in Sect. 73.3. We show our experiment in Sect. 73.4. At last, we conclude our approach in Sect. 73.5.

### 73.2 RP Sequence Scheme

In this section, RP sequence scheme is described, which is integration of region labeling and path sequence. We also present a query algorithm for our approach.

### 73.2.1 Description of RP Sequence

By the descriptions of positions of elements and string values on an XML tree, region label scheme of an XML document is given. As a classic query scheme, region labeling scheme can determine the ancestor–descendant and parent–child relationships between the nodes with containment property of the labels. Figure 73.4 shows the labeling result of Fig. 73.1.

We label last node of each path in the path sequence with the region labeling scheme. So, we give the definition of RP sequence scheme as follows:

**Definition 1** Region Path sequence

With the integration of the region labeling scheme and path sequence, Region Path scheme is a sequence of (*region label, path*) pairs:

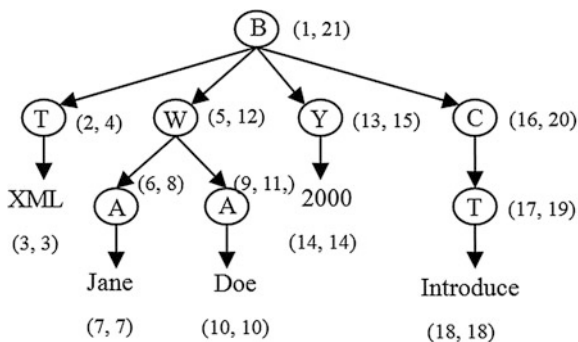
$$D = [(a_1, b_1, l), P_1][(a_2, b_2, l), P_2] \dots [(a_n, b_n, l), P_n]$$

where label  $(a_i, b_i, l)$  represents the label of the last node of path,  $p_i, l$  denotes id of the XML document and the sequence is in ascending order of  $a_i$ .

Similar to [8] and [5], we use capital letters to describe names of elements, and use a hash function to map the string values. For example,  $V_1 = h(\text{"XML"})$ ,  $V_2 = h(\text{"Jane"})$ . So, with RP sequence scheme, Fig. 73.1 can be represented as follows:  $D = [(1, 21, X), B][(2, 4, X), BT][(3, 3, X), BTV_1][(5, 12, X), BW][(6, 8, X), BWA][(7, 7, X), BWAV_2][(9, 11, X), BWA][(10, 10, X), BWAV_3][(13, 15, X), BY][(14, 14, X), BYV_4][(16, 20, X), BC][(17, 19, X), BCT][(18, 18, X), BCTV_5]$ , and  $X$  is id of the document. We also omit id of document in the label when there is a single document.

RP sequence scheme can correctly reflect the data distribution of the XML document. The precise tree structure can be gained directly from our sequence scheme, although there are the cases of identical sibling node [5].

**Fig. 73.4** Region labeling representation of Fig. 73.1



### 73.2.2 Avoiding False Alarm

The relationship between two nodes can be determined easily with the containment property of region label [6]. The property of *ancestor-descendant* is described as follows:

*In a single document, the node  $n_1$  is a descendant of the node  $n_2$  if and only if  $a_1 > a_2$  and  $b_1 < b_2$*

For example, in Fig. 73.4, the string “Jane” with label (7, 7) is a descendant of the node W with label (5, 12), and the string “XML” with label (3, 3) is a child of the node T with label (2, 4).

We describe an XML document with RP sequence scheme and a query with path sequence without containing the interval. With the containment property of region label, false alarm caused by identical sibling nodes can be completely avoided, because the relationship of two nodes is unique.

We can denote  $S1$ ,  $S2$  from Fig. 73.3 with RP sequence scheme and  $Q$  with path sequence as follows, and here we omit ids of documents

$$S1 = [(1, 141), P][(2, 7), PQ][(3, 43), PQT][(5, 63), PQS][(8, 13), PR][(9, 10) PRU][(11, 12), PRT]$$

$$S2 = [(1, 10), P][(2, 5), PQ][(3, 4), PQT][(6, 9), PQ][(7, 8), PQS]$$

$$Q = (P, PQ, PQT, PQS)$$

With the new sequence scheme, the false alarm can be avoided. According to the containment property of the region labeling, the interval of the parent of  $PQS$  does not contain the interval of  $PQT$ , so  $S2$  does not match  $Q$ .

## 73.3 Subsequence Matching Algorithm

In this section, the algorithm for subsequence matching is presented. An auxiliary structure such as suffix tree [8] or virtual trie structure [2, 5] need not be constructed, and the part of region labeling in the region labeling scheme is enough to decide the relationship between any two nodes. Labels in the region labeling scheme are used to avoid false alarm and to finish subsequence matching correctly.

### 73.3.1 Index Structure

Maintaining an in-memory index for the sequences is unsuitable, as the index size grows mainly with the total length of the sequences. In essence, we would like to build an efficient disk-based index. Similar to ViST and PRIX, we also maintain a

two level B<sup>+</sup>-tree to finish the subsequence matching. The structure is designed as follows.

- We first insert all the different paths in the sequences from documents into the B + -Tree using these paths as keys. We call this B + -Tree the **Path B<sup>+</sup>-Tree**.
- Each leaf node in the Path B<sup>+</sup>-Tree points to another B<sup>+</sup>-Tree called **Element B<sup>+</sup>-Tree**. This B<sup>+</sup>-Tree is indexed on the document IDs as keys and it stores the intervals from the Region Path sequence.

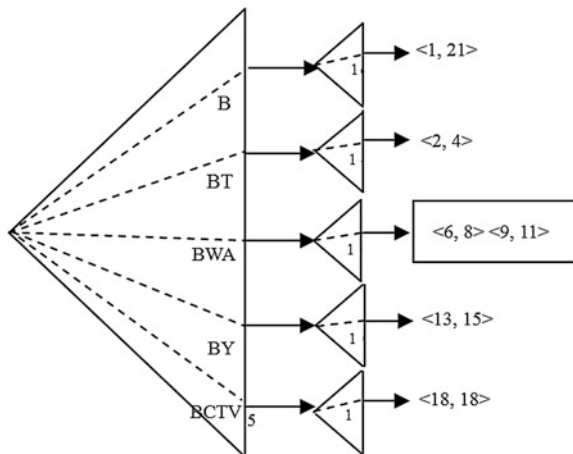
Figure 73.5 shows the two level B<sup>+</sup>-tree structure of Fig. 73.4

### 73.3.2 XML Data Subsequence Matching

Here, a brief description of how query processing is done by subsequence match is given through an example. Suppose we have the following query sequence  $(q_1, q_2, q_3, \dots, q_n)$ , where  $q_i$  is a path from the sequence and matches the node  $X$ . Firstly, by using path  $q_i$  as the key, we search a corresponding leaf node in Path B<sup>+</sup>-Tree structure, and return the label (s) of  $X$  from Element B<sup>+</sup>-Tree. Then, in Element B<sup>+</sup>-Tree corresponding to the path  $q_{i+1}$ , we find the descendant (s) of  $X$  with the containment property of region labeling. For each descendant, we repeatedly use the same process to match the next path in the query sequence above until we reach the last node of the query.

The above subsequence matching can be easily extended to multiple documents. The only difference is that after getting all the intervals in the first matching we get next matching according to documents IDs, and the documents ID must be identical. The following algorithm formalizes the process.

Fig. 73.5 Index structure



**Algorithm 1:** Subsequence Matching

**Input:** a query sequence  $Q = \{q_1, q_2, q_3, \dots, q_n\}$   
 Path B<sup>+</sup>-Tree, index of paths  
 Element B<sup>+</sup>-Tree, index of nodes with  $\langle \text{StartPos}, \text{EndPos} \rangle$  labels in XML documents

**Output:** all occurrences of  $Q$  in all the documents

```

/* Assume (a0, b0) is label of the root in XML document
*/
Query((a0, b0, x), 1);

Procedure Query ((ai, bi, x), i+1)
  if i < |Q| then /* |Q| denotes the number of the paths
                    in Q */
    I ← header link of qi+1;
    N ← nodes from I matching qi+1 within the range [ai,
        bi], and X must be identical;
    for each node n ∈ N do
      /* n is labeled with (ai+1, bi+1) */
      Query ((ai+1, bi+1, x), i+2);
    end
  end

```

With containment property of region labeling, subsequence matching algorithm avoids the false alarm. For example, in  $S_2$  in Fig. 73.2, the interval of “PQS” is beyond that of the first “PQ”, so the “PL” is not an ancestor of “PQS”, and we cannot match the structure with Query  $Q$ .

## 73.4 Experiment and Analysis

We demonstrate effectiveness of RP sequence scheme with extensive experiments. In this section, we describe these experiments and present the results.

### 73.4.1 Experiment Setup

Experiments are implemented in Java with JDK 1.5.0. We conduct the experiments on the AMD Athlon XP 1.83 G with 256 MB main memory running on Windows XP (sp2) with 80 G hard disk.

We carried out our experiments on real XML database DBLP [9] and synthetic dataset XMark [10].

DBLP is a popular computer science bibliography database and each record of DBLP accords to a publication. The document trees in the DBLP dataset have good similarity in structure. The segment we downloaded has size of 121 MB with tree structure of maximum depth 6.

XMark which simulates information about activities of an auction site is a single record with a very large and complicated tree structure. In our experiments, we use an XMark database with scaling factor 1.0, 110 MB of data.

### 73.4.2 Query Performance

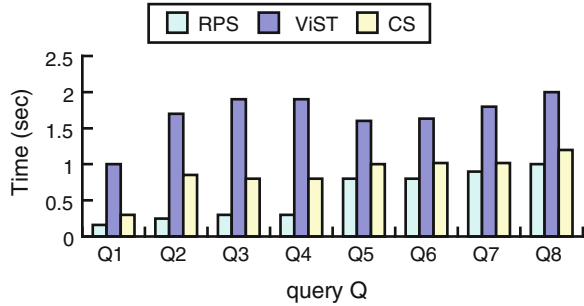
In order to compare the performance of RP sequence scheme (RPS) with that of ViST and Constraint Sequence, queries in Table 73.1 are executed in turn. Table 73.1 lists eight queries with ascending complexity in terms of selectivity, presence of values and twig structure. The first four queries from DBLP dataset represent four kinds of queries respectively and have different characteristics.  $Q_1$  is a single path expression query with two nodes and there is no attribute values involved.  $Q_2$  has one branch structure which involves an attribute value.  $Q_3$  and  $Q_4$  are twig patterns with three nodes and one branch respectively, and use wild cards which increase query scope. For XMark dataset, there are also four queries  $Q_5$ ,  $Q_6$ ,  $Q_7$  and  $Q_8$ .  $Q_5$  and  $Q_6$  are simple queries with four nodes and five nodes respectively. Both of them don't have any value.  $Q_7$  is a twig query with seven nodes and two branches.  $Q_8$  is a twig query with five nodes and one branch. Both  $Q_7$  and  $Q_8$  have wild cards and attribute values. In Fig. 73.6, we summarize the performance results in total time elapsed for the queries listed in Table 73.1.

Figure 73.6 shows the performance for the eight queries in RPS, ViST and CS. Firstly, we tested  $Q_1$ ,  $Q_2$ ,  $Q_3$ , and  $Q_4$  for DBLP dataset. The performance difference is due to the following reasons. First, similarity in the sequences from DBLP dataset is very high, which maybe produce many Identical Sibling Nodes and cause the false alarm problem. ViST's sequencing and query algorithms do not guarantee query equivalence between a structure match and a subsequence match, and expensive join operations [2] is used to avoid false alarm, which takes much more time. In CS, in order to void false alarm and maintain query equivalence, the notion of constraint is introduced, which makes query performance of CS better than that of ViST. In addition, CS takes advantage of sharing of sequences in case of simple path expression which doesn't contain any value, so the performance for

**Table 73.1** Sample queries over DBLP dataset

	Path Expressions	Datasets
$Q_1$	/mastersthesis/title	DBLP
$Q_2$	/article[mdate = "2003-01-31"]/title	DBLP
$Q_3$	*/author[text = "Peter"]	DBLP
$Q_4$	//author[text = "Peter"]	DBLP
$Q_5$	/site/people/person/name	XMark
$Q_6$	/site/regions/africa/item/reserve	XMark
$Q_7$	/site/item[quantity = "2"][[location = "United States"]]/name	XMark
$Q_8$	/site/open_auctions/open_auction[//author/person = "person11080"]/time	XMark

**Fig. 73.6** Comparison of queries processing times



query  $Q_1$  is much better than that of our approach. Tree labeling in CS doesn't maintain query equivalence, and the constraint is used to decide the relationship between two nodes, which costs more time than our approach. Second, in ViST and CS, an auxiliary structure that labels every element in the sequence is constructed to finish subsequence match. In the RP sequence scheme, the relationship between two nodes can be decided by intervals, so we need not spend extra time on an auxiliary structure.

For the XMark dataset, queries  $Q_5$ ,  $Q_6$ ,  $Q_7$  and  $Q_8$  are used to test query performance of PR, ViST and CS. Since XMark has deep and complicated structure, sharing of sequence is less, which makes identical sibling nodes less than that in DBLP. Figure 73.6 shows the result of comparison.  $Q_5$  and  $Q_6$  are simple path expressions without any value. Because there is less sharing of sequence, CS also shows good query performance.  $Q_7$  and  $Q_8$  are more complicated, which have branches, values and wildcards. Because performance-oriented strategy which reduces research space is used in CS, the performance comes closest to that of RP sequence. In Fig. 73.6, the performance of ViST for  $Q_5$ ,  $Q_6$ ,  $Q_7$  and  $Q_8$  is also described, and that is worst.

### 73.5 Conclusion

In this paper, we introduce an XML index approach based on sequence scheme, which also processes query holistically, without breaking the twig into many individual root-to-leaf paths, and avoids useless intermediate results and expensive join operations. Compared with previous index sequence approach, such as ViST, PRIX, Constraint Sequence, we can process the problem caused by query non-equivalence more efficiently. Combining with the region labeling scheme, we can always get precise resultant structure only by containment property of the labels. Experiment results show RP sequence scheme outperforms previous approaches on query performance.



**Acknowledgments** This work was supported by Science and Technology Foundation Platform Construction Project of Shanxi Province of China (Project No. 2012091003-0101) and partially supported by the Youth Fund of Taiyuan University of Technology (No. k201032).

## References

1. Bray T, Paoli J, Sperberg-McQueen C (1998) Extensible markup language (XML) 1.0. W3C recommendation available at <http://www.w3.org/TR/1998/REC-xml-19980210>
2. Raw PR, Moon B (2004) PRIX Indexing and querying XML using prüfer sequences. ICDE 2004, pp 288–300
3. Hilma K, Sreenivasa P (2005) Efficient indexing and querying of XML data using modified prüfer sequences. CIKM 2005, pp 397–404
4. Hilma K, Rajesh C, Sreenivasa P (2005) Handling updates in sequence based XML query processing. COMAD 2005
5. Wang H, Meng X (2005) On the sequencing of tree structures for XML indexing. ICDE 2005, pp 372–383
6. Srivastava D, Al-Khalifa S, Jagadish HV (2002) Structural joins: a primitive for efficient XML query pattern matching. ICDE 2002, pp 141–152
7. Boag S, Chamberlin D, Fernandez MF, Florescu D, Robie J, Siméon J (2002) XQuery 1.0: an XML query language. <http://www.w3.org/TR/xquery>
8. Wang H, Park S, Fan W, Yu PS (2003) ViST: a dynamic index method for querying XML data by tree structures. In proceedings of SIGMOD 2003, pp 101–121
9. DBLP database web site. <http://uni-trier.de/XML>
10. Schmidt A, Wass F, Kersten M, Carey J, Manolescu I, Busse R (2002) XMark: a benchmark for XML data management. Proceedings of the 28th international conference on very large databases 2002

# Chapter 74

## Model-Based Workpiece Positioning for Robotic Fixtureless Assembly Using Parallel Monocular Vision System

Weiwei Yu, Mingmin Zhai and Yasheng Chen

**Abstract** This paper proposed to use parallel monocular vision system that could fit different robotic grasping pattern, in order to reduce the computation burden for real-time grasping control. A novel model-based workpiece positioning approach, which can solve both 3D or 2D pose estimation problem, is proposed by using the imagery template and homography matrix. The demand of workpiece template is not 3D model, but the workpiece template image, which is much easier to obtain. Moreover, as the positioning expressions based on homography matrix between the workpiece template and images, and the two camera images, are expressed as a simple formula, the proposed approach is intuitive for algorithm development.

**Keywords** Workpiece positioning · Robot fixtureless assembly · Parallel monocular vision · Model-based pose estimation

### 74.1 Introduction

Robotic Fixtureless Assembly, that eliminates the need of using complex and rigid fixtures, is a challenging task which requires solving many innovation problems in robotics, including precision control, intelligent grasping, sensor integration, part mating etc. [1]. In particular, the development of a robust, and efficient machine vision system empowers the robot to identify and determine the pose of the workpiece. With the help of such vision system, the robot could recognize, servo to and grasp the target part which may placed with other workpieces in an unstructured manner, thus enable the automatic robotic assembly.

---

W. Yu (✉) · M. Zhai · Y. Chen  
School of Mechanical Engineering, Northwestern Polytechnical University, 228#,  
Xi'an 710072 Shaanxi, China  
e-mail: yuweiwei@nwpu.edu.cn

According to the location of camera in the robotic fixtureless assembly, it can be divided into eye-in-hand and eye-on-head design [2]. For the case of eye-on-head design, the 3D invariant object recognition and pose calculation can be achieved by using 2D computer vision in different manner. As the common way of positioning 3D ordinary object with images, the 3D workpiece pose estimation can be achieved by the reconstruction through multi-view epipolar geometry [3]. Because of the complex algorithm and time consuming of reconstruction process, it hardly satisfy the real-time requirement of robot vision servo system.

In recent years, more and more intelligent algorithms, such as Genetic Algorithm, and Neural Network are introduced to the workpiece positioning area. Feature CMAC neural network is proposed for pose estimation for fixtureless assembly in [4]. This method has generalization ability and do not rely on the specific mathematic model of workpiece. A Fuzzy Adaptive Resonance Theory is applied for 3D invariant object recognition and pose calculation for aligning parts in assembly task [5]. Besides the intelligent algorithm based method can satisfy the real-time requirement well, it needs a long period training time priority.

Regardless the reconstruction of object, it is usually to use the affine transformation to estimate the 3D workpiece position from the 2D images. One branch is upon the weak perspective projection. Wei Liu proposed a method of workpiece positioning from perspective view based on Differential Evolution algorithm [6]. Another based on the model-based pose estimation method which usually indicates the shape, texture or the appearance of the object in advance. In [7], a general method for fitting models with arbitrary curved surfaces and any number of internal parameters to matched image features is introduced. Chan-Ho Lee introduced the virtual plane method for the 3D pose estimation of large objects using single camera systems [8]. As the virtual plane method used multiple sets of single camera and the CAD information of the object, this method required more straightforward calculation than conventional solutions.

As the model-based pose estimation is based on the prior knowledge of target object, the stable matches between the template and image data make it reliable to estimate the pose from 2D image data. Moreover, with the help of the model, this method required simpler analysis than other solutions [8]. However, if considering the robotic grasping pattern during the assembly, workpiece positioning can be divided into two cases. In the first case, its pose may influence the robotic grasping or aligning performance, the 3D position of the workpiece center and its rotation should be determined for the robot to choose the optimal grasping way. In the second case, if the working part is restricted with sliding on the workbench with always the same face up, the problem is now simplified as 2D positioning of the workpiece center and its rotation.

Therefore, the objective of this research is to design a feasible method that could fit both of the above two cases for eliminating the image analysis complexity and decreasing processing time for on-line control at the same time. A model-based approach is presented, in which the workpiece pose can be directly calculated by the target template image and some homography matrix.

### 74.2 Parallel Monocular Vision System

The workpiece pose estimation problem can be simplified into two cases. If consider the grasping pattern which may hamper robotic performance for fixtureless assembly and control complexity, the problem can be simplified as finding  $X - Y - Z$  position of the workpiece center and its rotation about  $Z$  axis. However, if one can reasonably restrict the working parts sliding on the worktable without seriously hampering the ability to accomplish the assembly, for example, imagine a human operator casually placing parts on the table, provided the desired face is always up, the problem now becomes finding the  $X - Y$  position of the workpiece center and its rotation around  $Z$  axis [4].

In the first case, because the depth information of the target part should be calculated, a stereo camera is always chosen. If only  $X - Y$  position of the workpiece is needed for grasping, a monocular camera is enough. The eliminating of the image matching time could increase the vision servo system on-time efficiency. Therefore, we propose to use two monocular cameras of the same type whose optical axis are installed strictly parallel. The monocular and stereo vision can be switched automatically according to different requirement. The installation and system module of parallel monocular vision is represented in Fig. 74.1.

### 74.3 Workpiece Positioning with Parallel Monocular Vision

Consider the location of gravity center of object in the workbench coordinate system as the pose of the workpiece, since the workbench whose coordinate is easy transformed to robot coordinate, is always fixed. And its orientation is determined by the direction angle of the envelope rectangle of the workpiece. The workpiece model data base contains three type of information: workpiece template image, its keypoints and descriptor, and the gravity center pixels of the template image.

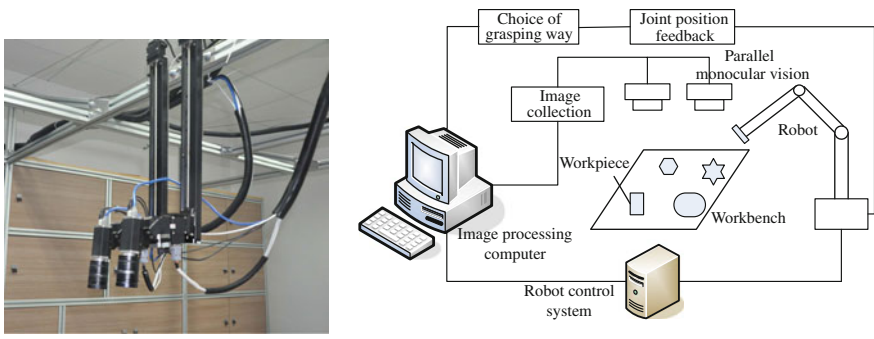


Fig. 74.1 Installation and system module of parallel monocular vision

Different from other model-based methods, the model required in our experiment is only template image, but not the 3D object model.

### ***74.3.1 Workpiece Positioning with Two Parallel Monocular Cameras***

In the first case with two parallel monocular cameras, the scheme contains three stages of image matching: (1) feature matching between the left and right camera image, (2) left camera image matching with the target workpiece template, (3) right camera image matching with the template. The SIFT algorithm is chosen for all the matching process, for its advantages of keypoints shown to be invariant to image rotation and scale, and robust of affine distortion and change in illumination. In order to reduce the running time and increasing the robustness, the image background subtraction and segmentation is adopted before first matching step. The workpiece position can be intuitive calculated by the projection equation, homography matrixes between the right and left camera images with the target workpiece template image.

### ***74.3.2 Workpiece Positioning with One Monocular Camera***

In the case of one monocular camera, the orthography of the workbench with markers is taken as the workbench template. Different from the first case, in this scheme, there are only two image matching process should be included: one is the matching between workbench template and one of the camera image (take the left camera image as example), the other is between the left camera image and the target workpiece template. By the first homography matrix, the workpiece image is transformed to orthographical in order to estimate the object orientation. The workpiece position can be calculated based on the homography matrix between the target object image and workpiece template image.

Comparing the two workpiece positioning process, just by substitute the matching process between the left and right image, right image and workpiece template with the matching between the left image and workbench template, it is easy to achieve one scheme that could fit both of the two cases.

## **74.4 Workpiece Position Calculation Based on Homography Matrix**

The principle of object positioning by two parallel monocular cameras is as same as binocular vision. Assume that the coordinates of arbitrary object point in the left camera coordinate system is  $(x_l, y_l, z_l)$ , in the left and right image the coordinates

are  $(u_l, v_l)$  and  $(u_r, v_r)$  accordingly. Based on triangulation principle, the coordinates in the left camera coordinate system are:

$$x_l = \frac{z_l(u_l - u_o)}{\alpha_x} \quad (74.1)$$

$$y_l = \frac{z_l(v_l - v_o)}{\alpha_y} \quad (74.2)$$

$$z_l = \frac{b\alpha_x}{u_l - u_r} \quad (74.3)$$

In the above equations,  $\alpha_x$ ,  $\alpha_y$ ,  $u_o$  and  $v_o$  are the camera intrinsic parameters.  $\alpha_x$  and  $\alpha_y$  are the camera scale factor along  $x$  and  $y$  axis, and  $u_o$ ,  $v_o$  are the camera principal point coordinates.  $b$  is distance between the camera projection centers.

For a given plane in world coordinate system, the homography matrix represents the relationship of every feature belonging to the plane in one image and the corresponding feature in another image from different angle. Assume that

$$p = (u_t, v_t) \quad (74.4)$$

$$q = (u_l, v_l) \quad (74.5)$$

are the corresponding feature point coordinates from the workpiece template image and the left monocular camera image respectively. There exists a  $3 \times 3$

homography matrix  $H = \begin{bmatrix} h_0 & h_1 & h_2 \\ h_3 & h_4 & h_5 \\ h_6 & h_7 & h_8 \end{bmatrix}$ , such that

$$q = Hp \quad (74.6)$$

RANSAC is adopted to reject outliers and find the homography matrix between workpiece templates and current image. Suppose that  $p$  and  $q$  represent the two dimension pixel coordinates of workpiece center. For affine transform  $h_8 = 1$ , Eq. (74.6) can be written as

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & h_2 \\ h_3 & h_4 & h_5 \\ h_6 & h_7 & 1 \end{bmatrix} \begin{bmatrix} u_t \\ v_t \\ 1 \end{bmatrix} \quad (74.7)$$

Thus, the workpiece gravity center of the left monocular camera image  $(u_l, v_l)$  can be expressed by the position of gravity center of target workpiece template  $(u_t, v_t)$ :

$$u_l = (h_0u_t + h_1v_t + h_2)/(h_6u_t + h_7v_t + 1) \quad (74.8)$$

$$v_l = (h_3u_t + h_4v_t + h_5)/(h_6u_t + h_7v_t + 1) \quad (74.9)$$

In the monocular camera case, the  $X - Y$  position of the gravity center of object can be calculated according to Eqs. (74.8) and (74.9), only related to the homography matrix and the gravity center pixels of template image.

If  $H' = \begin{bmatrix} h'_0 & h'_1 & h'_2 \\ h'_3 & h'_4 & h'_5 \\ h'_6 & h'_7 & h'_8 \end{bmatrix}$  represents the homography matrix between the

workpiece template image and right monocular camera image, with the same method, one can calculate the position of object gravity center as:

$$u_r = (h'_0 u_t + h'_1 v_t + h'_2) / (h'_6 u_t + h'_7 v_t + 1) \quad (74.10)$$

$$v_r = (h'_3 u_t + h'_4 v_t + h'_5) / (h'_6 u_t + h'_7 v_t + 1) \quad (74.11)$$

Assume the 3D gravity center of the workpiece is  $(C_{x_l}, C_{y_l}, C_{z_l})$  according to the projection relation Eqs. (74.1)–(74.3), substitute expression of gravity center of the left camera image which represented by the homography matrix with the object template respectively,

$$C_{x_l} = \frac{b[(h_0 - h_6 u_0)u_t + (h_1 - h_7 u_0)v_t + h_2 - u_o]}{(h_0 - h_6 u_r)u_t + (h_1 - h_7 u_r)v_t + h_2 - u_r} \quad (74.12)$$

$$C_{y_l} = \frac{b\alpha_x[(h_3 - h_6 v_0)u_t + (h_4 - h_7 v_0)v_t + h_5 - v_o]}{\alpha_y[(h_0 - h_6 u_r)u_t + (h_1 - h_7 u_r)v_t + h_2 - u_r]} \quad (74.13)$$

$$C_{z_l} = \frac{b\alpha_x(h_6 u_t + h_7 v_t + 1)}{(h_0 - h_6 u_r)u_t + (h_1 - h_7 u_r)v_t + h_2 - u_r} \quad (74.14)$$

Again, substitute position of object gravity center in the right camera image  $u_r$  in the above equation with expression (74.10), the pose of workpiece can be directly calculated with homography matrix between the workpiece template image and each of monocular camera image  $H, H'$ , the position of gravity center in target workpiece template image  $(u_t, v_t)$ , and camera intrinsic parameter  $u_o$ .

## 74.5 Experiment Results and Analysis

In the one monocular camera case mentioned in Sect. 74.2, the  $Z$  position does not hampered the robot grasping during the assembly. According to the workpiece positioning process in Fig. 74.2, take the orthography of the workbench with 4 markers at each corner as its template (Fig. 74.3a). After the feature matching between the workpiece template and the image, the image is transformed to orthography relative to workbench, in order to obtain the workpiece's orientation. The position and orientation of the object is calculated based on Eqs. (74.8) and (74.9), and results are indicated in Fig. 74.3b.

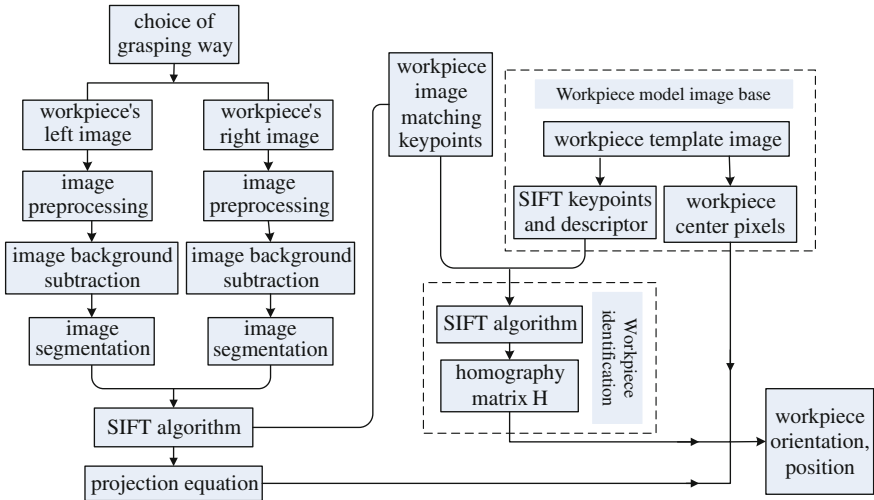


Fig. 74.2 Process scheme of workpiece positioning with two parallel monocular cameras

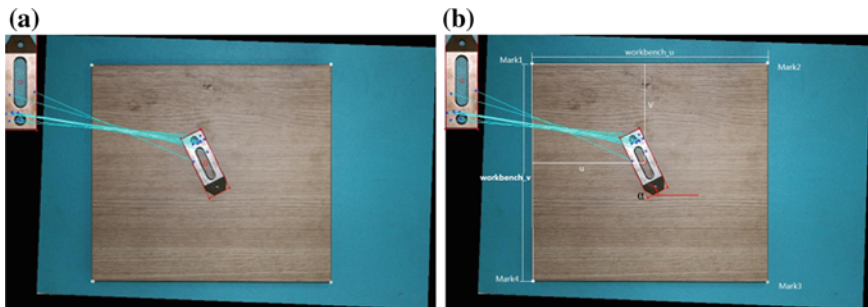


Fig. 74.3 Workpiece positioning with one monocular camera, **a** Image transform after feature matching. **b** Pose estimation

If consider the grasping way of robot, use two parallel monocular cameras to determine the  $X - Y - Z$  position for workpiece according to the process scheme in Fig. 74.4. After the background subtraction and morphological segmentation process, the object target identification and the feature matching with the workpiece template in the multi-object case, is shown in Fig. 74.5a. In Fig. 74.5b, the target object is half sheltered from the shadow of robot. It also shows good results by our proposed approach, because with the aid of morphological segmentation for identify overlapping objects. It should be emphasized that in both of these two cases, the images are in the situation of scaling and rotating.

If the difference between position estimation for the workpiece along any of  $X$ ,  $Y$  or  $Z$  direction and its real pose is smaller than the error threshold value, it is



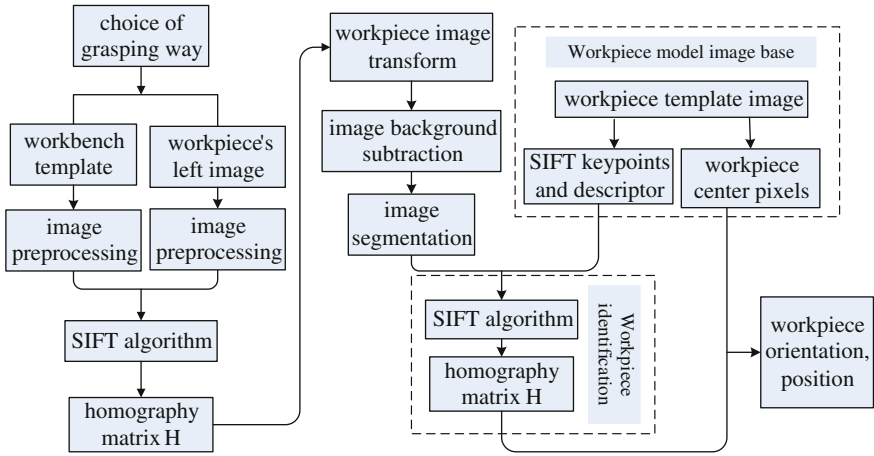


Fig. 74.4 Process of workpiece positioning with one monocular camera

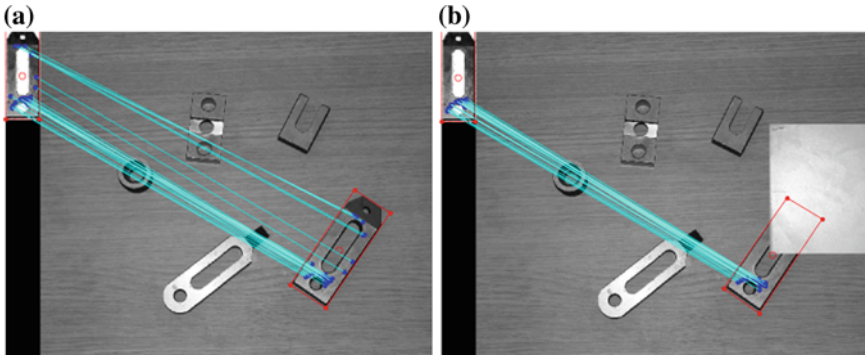


Fig. 74.5 Target object identification in the case of multi-object and obscured, **a** Feature matching in the multi-object case. **b** Feature matching in the obscured case

considered to be a success calculation. In this experiment, the error threshold is set to be  $3\text{ mm}$ , 80 pair different images are taken as examples, and the image pixel is  $1568 * 2352$ . As the relative position of the camera and object, all the images taken in this experiment are scaled and rotated. The calculated success rate in this case is 93.75%.

If substitute the SIFT algorithm for the workpiece image matching process in the proposed approach with the Harris corner detection, try these two methods in three situations: (1)the image without scaling and rotation, (2) image with obvious scaling and rotation, and (3)object with obscured. The comparing results are shown in Tables 74.1 and 74.2 in the case of  $1568 * 2352$  pixel image and  $480 * 640$  image respectively. It is apparently that adopted SIFT algorithm for the feature matching in this approach gets better results than the Harris corner detection in all the three situations. However, the success rate for both of the two

**Table 74.1** The success rate with two methods for 80 1568 \* 2352 images

	Without scaling and rotation (%)	With scaling and rotation (%)	With obscured (%)
Method with SIFT	96.25	93.75	95.00
Method with Harris	88.75	82.50	81.25

**Table 74.2** The success rate with two methods for 80 480 \* 640 images

	Without scaling and rotation (%)	With scaling and rotation (%)	With obscured (%)
Method with SIFT	93.33	90.00	91.25
Method with Harris	87.50	80.00	78.75

**Table 74.3** The success rate and running time with two methods for 80 1568 \* 2352 images

	Without background subtraction and morphological based segmentation (%)	With background subtraction and morphological based segmentation (%)
Success rate	83.33	93.75
Running time	100	84.3

methods is influenced by the image resolution, especially in the scaling, rotation and obscured cases.

Try to calculate the workpiece position without the step of background subtraction and morphological transformation based segmentation, and assume that the running time for this case is 100 %. Again, comparing the success rate and running time with our proposed method. According to the results listed in Table 74.3, the running time of the proposed approach is only 84.3 % of the previous method, and the success rate increased from 83.33 to 93.75 % visibly. That is because of with the help of the background subtraction and morphological based segmentation, the matching range is much smaller and more specific, the overall calculating time is much less. Moreover, the method is more robustness, since the segmentation result is independent of the shape or placement of the zones of interest.

## 74.6 Conclusion

Because of the requirement of different robotic grasping pattern, workpiece positioning for robotic fixtureless assembly can be simplified into 3D or 2D pose estimation problem. In order to eliminating the image processing complexity and decreasing time for on-line control at the same time, a parallel monocular vision system is presented. A novel approach is proposed to use the imagery target template and homography matrix to positioning the workpiece for robot grasping. Since

the positioning expressions based on homography matrix between the workpiece template and image and the two camera images, are expressed as a simple formula, the simulation results demonstrate the robustness and less time-consuming of the proposed approach which required simpler analysis. Because the 3D workpiece tracking problem appears often in recent robot assembly, we would like to introduce this approach to dynamic tracking of industrial part in future work.

**Acknowledgments** This research has been supported by the “111 Project (Grant No.B13044)”, and the fundamental research fund of NWPU (JC20100211).

## References

1. Corona-Castuera J, Rios-Cabrera R, Lopez-Juarez I, Peña-Cabrera M (2005) An approach for intelligent fixtureless assembly: issues and experiments advances in artificial intelligence. *Lect Notes Comput Sci* 3789:1052–1061
2. Song K-T, Chang C-H (2011) Object pose estimation for grasping based on robust center point detection. In: 8th Asian control conference, Kaohsiung, Taiwan, pp 305–310
3. Sengupta S (2009) Issues in 3D reconstruction from multiple views. Indian Institute of Technology Madras, Master thesis
4. Langley CS D’Eleuterio GMT (2000) Pose estimation for fixtureless assembly using a feature CMAC neural network. In: 31st International symposium on robotics, Montreal, Canada, pp 356–361
5. Peña M, López I, Osorio R (2006) Invariant object recognition robot vision system for assembly. In: Electronics, robotics and automotive mechanics conference, Cuernavaca
6. Liu W, Chen T, Wang P, Qiao H (2012) Pose estimation for 3D workpiece grasping in industrial environment based on evolutionary algorithm. *J Intell Robot Syst* 68:293–306
7. Lowe DG (1991) Fitting parameterized three-dimensional models to images. *IEEE Tans Pattern Anal Mach Intell* 13(5):441–450
8. Lee C-H, Oh J-K, Lee S-H et al (2010) Development of a robot vision system for measuring 3D pose of large object using virtual plane algorithm. In: 41st international symposium on robotics, Muich, Germany, pp 340–345

# Chapter 75

## Single-Trial Identification of Motor Imagery EEG based on HHT and SVM

Peng Lu, Daoren Yuan, Yafei Lou, Chi Liu and Shilei Huang

**Abstract** Single-trial identification of motor imagery (MI) EEG is one of the key techniques in the brain-computer interface (BCI). To improve the accuracy of classification and reduce the algorithm time, targeting at motor imagery (MI) EEG of four kinds of motion, a single-trial identification algorithm of MI EEG based on HHT and SVM is proposed. Firstly, MI EEG is decomposed into 8-order intrinsic mode function (IMF) and margin R by empirical mode decomposition (EMD). Secondly, Hilbert spectrum is got by Hilbert transformation. AR model parameter of the extracted 6-order IMF is extracted. The acquired 6-order AR parameter and the characteristic quantity of 29 power spectral density included in the 4-32 Hz EEGs constitute a 35 dimensional characteristic vector. Finally, support vector machine (SVM) is used to classify. The single-trial identification results are as follows: the average recognition rate of the two kinds of thinking actions is 91.6478 %, and that of three is 89.4798 %, four is 89.4064 %.

**Keywords** Motor imagery · HHT · SVM · Single-trial identification

### 75.1 Introduction

Brain-computer interface (BCI) technique is one of the hottest research spots at present. Evoked EEG based on SSVEP and spontaneous EEG based on MI are two important means to study BCI [1, 2]. The former requires a structured environment with special apparatus, while the latter is a man-machine interface mode of independent, non-dependent, non-structured environment [3, 4]. At present, problems like long training cycle, long consuming of algorithm and low rate of the

---

P. Lu (✉) · D. Yuan · Y. Lou · C. Liu · S. Huang  
School of Electrical Engineering, Zhengzhou University, No.100 Science Road, Zhengzhou, China  
e-mail: lupeng@zzu.edu.cn

thinking action classification challenge the online BCI system [5]. Therefore, shortening the training cycle, lowering the algorithm complexity and improving the rate of thinking action identification are the key techniques for online BCI.

The single-trial identification of MI EEG includes signal pretreatment, feature extraction, thinking action identification and so on. And feature extraction becomes the core technology of BCI. The method of feature extraction, which is the core and key technology, includes WT [6], CSP [7], ICA [8] et al. For the nonlinear and non-stable MI EEG, the above-mentioned methods exit the problems of worse self-adaptive, more amounts of channels, non-independent component.

To improve the accuracy of classification and reduce the algorithm time, a single-trial identification algorithm of MI EEG based on HHT and SVM, targeting at MI EEG of four kinds of motion, has been devised to address this issue. Firstly, MI EEG is decomposed into 8-order intrinsic mode function (IMF) by empirical mode decomposition (EMD). Secondly, Hilbert spectrum is got by Hilbert transformation. AR model parameter of the extracted 6-order IMF is extracted. The acquired 6-order AR parameter and the characteristic quantity of 29 power spectral density included in the 4-32 Hz EEG constitute a 35 dimensional characteristic vector. At last, Support Vector Machine (SVM) is used to classify. The single-trial identification results are as follows: The average recognition rate of the two MI tasks is 91.6478 %, and that of three is 89.4798 %, four is 89.4064 %.

## 75.2 Data Acquisition

The data comes from the non-implanted brain-computer interface laboratory of Zhengzhou University. The MI EEG data is adopted by a 64-channel 10–20 system Quik-Cap.

**Subjects:** The experimental objective is explained to the subjects in detail. Four healthy subjects (three males and one female, the average age 24.5 years old) participated in the EEG data acquisition. All the subjects are dextrorality. No subjects have known sensory-motor disease or psychological medical history.

**Experimental paradigm:** The data recording process for each time is as follows: at the beginning of the experiment ( $t = 0$  s), “+” appears in the black screen, when  $t = 1$  s, “+→” appears. The left, right, upward and downward direction arrows (corresponding to the left hand, right hand, left foot and right foot in the four kinds of motor imageries) are truly displayed when  $t = 2$  s with duration 4 s. The task was asked to do the corresponding imagery motor according to the direction of the arrow. When the black screen presents, a temporary rest of 1 s was taken. Each experiment contains two rounds made up of 40 trails, which has a five minutes rest.

**Data acquisition:** The MI EEG primary data from C3, C4, FC3 and FC4 was analyzed by the relevance analysis.

### 75.3 Algorithm

#### 75.3.1 Hilbert-Huang Transforms

Hilbert-Huang transforms (HHT) is a stable processing process of non-stable and nonlinear MI EEG [9], gradually decomposing the trend terms of different scale fluctuation and signal contained in EEG through the iterative process of EMD, then decomposing the MI EEG into the sum of IMF in different frequencies as Fig. 75.1.

HHT, having high time–frequency characteristic, is divided into two parts: EMD and Hilbot transform. At first, any signal  $s(t)$  is decomposed into the sum of IMFs by EMD.

$$s(t) = \sum_{i=1}^N c_i(t) + r_N(t) \tag{75.1}$$

Then spectrum analysis is performed by Hilbert transform for every IMF component. The marginal spectrum can be further defined by Hilbert spectrum as

$$h(w) = \int_{-\infty}^{+\infty} H(w, t)d(t) \tag{75.2}$$

#### 75.3.2 Support Vector Machine

The main idea of support vector machine (SVM) algorithm [10] is mapping the input vector to a high-dimensional feature space using kernel function and constructing the optimal separating hyper plane in this space which would be finally

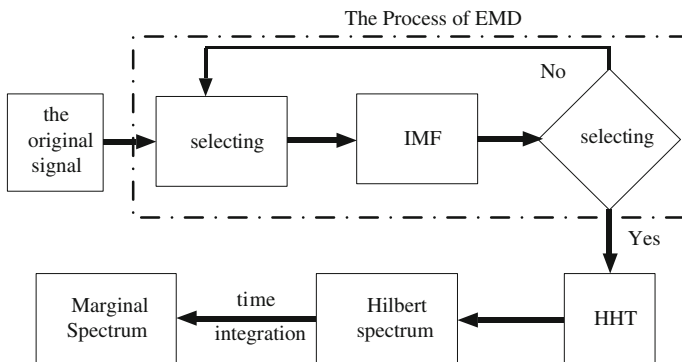


Fig. 75.1 The process of HHT

coming down to the convex quadratic programming problem. The algorithm can be described as:

- (1) Setting the known training set

$$T = \{(x_i, y_j)\}, i = 1, 2, \dots, l, \text{ where } x_i \in R^n, y_i \in \{1, -1\};$$

- (2) Constructing and solving optimization problem:

$$\left\{ \begin{array}{l} \min \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j a_i a_j (x_i x_j) - \sum_{j=1}^l a_j \\ \text{s.t. } \sum_{i=1}^l y_i a_i = 0 \\ a_i \geq 0, \quad i = 1, 2, \dots, l \end{array} \right. \quad (75.3)$$

Get the optimal solution  $a^* = (a_1^*, \dots, a_l^*)^T$ ;

- (3) Calculating  $w^* = \sum_{i=1}^l y_i a_i^* x_i$ ;

A positive component  $a_j^*$  of  $a^*$  is selected to calculate  $b^* = y_j - \sum_{i=1}^l y_i a_i^* (x_i x_j)$

- (4) The optimal separating hyper plane  $w^* x + b^* = 0$  is constructed, and the decision function is got,  $f(x) = \text{sgn}(w^* x + b^*)$ .

## 75.4 Feature Extraction

### 75.4.1 Pretreatment

Coherent averaging method is an important single-trial EEG extraction method in electrophysiological research [11]. The recorded observational signal  $x(t)$  is composed by practical EEG  $s(t)$  and noise, and noise is usually stronger than response. Therefore, we repeat this stimulus related experiments several times and got the cumulative average of several observe results. Each record was set as:  $x_i(t) = s_i(t) + n_i(t)$ ,  $i = 1, 2, \dots, N$ .

The starting time of each record is the instant of imposing stimulation. Accumulation should be performed in the same starting point, which is called coherent average.  $\bar{x}(t) = \frac{1}{N} \sum_{i=1}^N x_i(t)$  is called average evoke.

It can be seen that coherent averaging method is a simple and effective single-trial extraction method in Fig. 75.2.

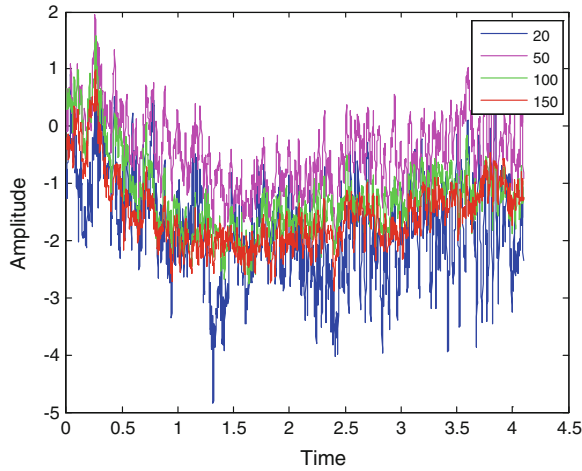


Fig. 75.2 EEG superimposed in different times of left hand MI in C4

### 75.4.2 Feature Extraction

The analysis data of EGG signal of left hand MI in C4 channel are taken as an example. At first, 8-order IMF components are got by EMD, which is demonstrated in Fig. 75.3.

From AR power spectrum analysis, the component range of effective EEG and mode component of each rhythm component are determined. It indicates that the frequency range of the first mode component is centralized in above 40 Hz in Fig. 75.4; obviously, this component is mainly high-frequency noise.

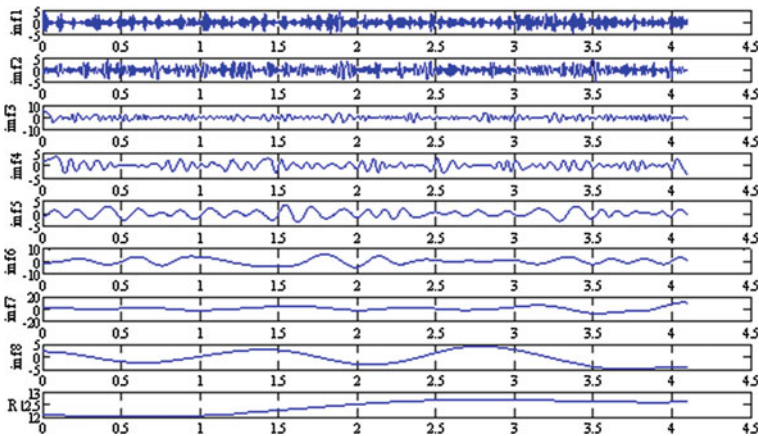


Fig. 75.3 EMD of left hand MI in C4 channel



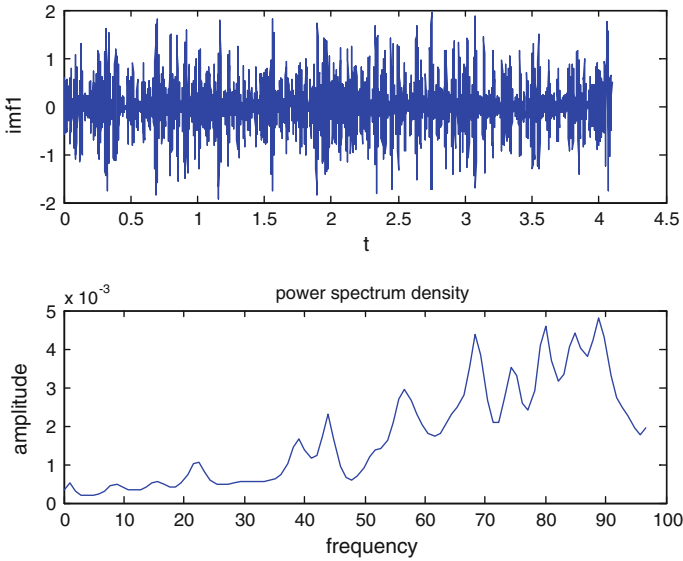


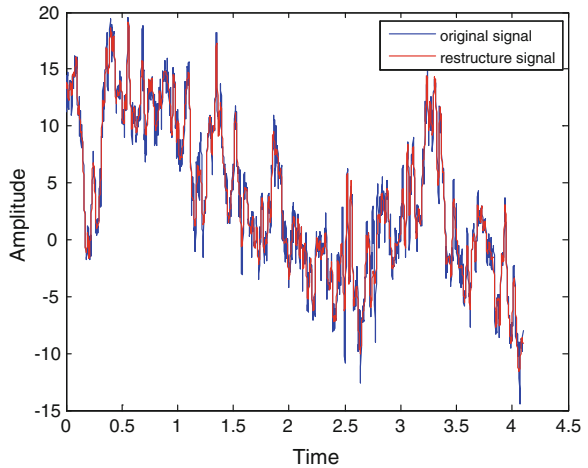
Fig. 75.4 IMF1 and its power spectral density

From the above analysis, we can see the effective EEG component is centralized in 3–8 mode components. They were superposed and reconstructed. The comparison of reconstructed EEG and original signal is demonstrated in Fig. 75.5.

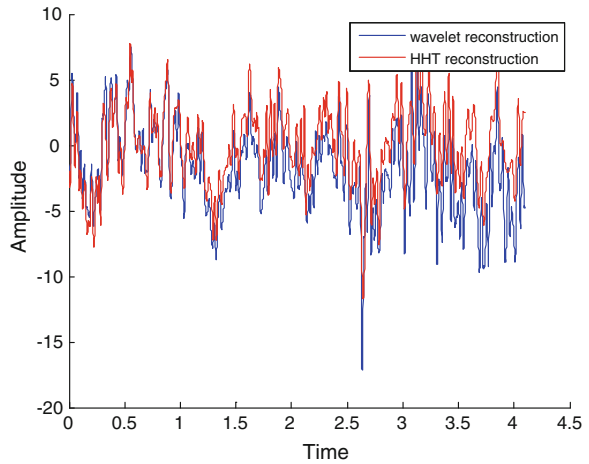
HHT and WT method were respectively used to do single-trial EEG extraction for the same data of left hand EEG in C4 channel, and the EEG extraction efficacy was compared.

HHT and WT were respectively used to do single-trial EEG extraction for the left hand EEG in C4 channel. The extraction results were shown in Fig. 75.6. Both

Fig. 75.5 Comparison chart



**Fig. 75.6** Comparison of two methods



reconstructions can effectively express the MI EEG, but HHT reconstruction is evidently smoother than WT because the primary function of WT is predefined, yet HHT is depended on EEG itself, having the self-adaptional decomposition and reconstitution in time domain.

## 75.5 Classification

8-order IMFs which were got after HHT and AR model parameter was extracted from the last 6 order IMFs. The power spectral density of 4–32 Hz EEG is regarded as the characteristic.

The characteristic vector of single channel EEG includes 6 model parameters and 29 power spectral densities, which would synthesize a 35-dimensional characteristic vector.

SVM was used for classification. 600 trails are the training data of the four kinds of motor patterns, 143 trails are regarded as the test data. The classification results of C3, C4, FC3 and FC4 were tested by multiple SVM classifiers.

In order to verify recognition accuracy and real-time performance of the proposed method, five experiments were done, and the experimental results are shown in Tables 75.1, 75.2, 75.3, 75.4, 75.5. In order to verify the real-time of online BCI, three methods are used for the single-trial and identification of the collected 200 groups of primary EEGs. The comparison of the running time is demonstrated in Table 75.5.

Above five experimental groups, these can be obtained as follows: the average single-trial identification rate based on HHT and SVM is 91.6478, 89.4798 and 89.4064 %, which is more than 70 % higher than that in literature [12]; the first three tables indicated that using the algorithm of this study high classification accuracy has been got from MI EEG, and with the increase of imagery motion, the

**Table 75.1** Identification results of two kinds of motions

Motion category				Identification accuracy (%)
Left hand	Right hand	–	–	92.4731
Left hand	–	–	Right foot	89.1720
–	Right hand	Left foot	–	91.3978
–	Right hand	–	Right foot	93.5484
Average identification accuracy				91.6478

**Table 75.2** Identification results of three kinds of motions

Motion category				Identification accuracy (%)
Left hand	Right hand	Left foot	–	91.2032
Left hand	Right hand	–	Right foot	89.0125
Left hand	–	Left foot	Right foot	89.5305
–	Right hand	Left foot	Right foot	88.1729
Average identification accuracy				89.4798

**Table 75.3** Identification results of four kinds of motions in single channel

Lead channel	Identification accuracy (%)
C3	90.2098
C4	88.1119
FC3	89.5105
FC4	88.1119
Average identification accuracy	89.4064

**Table 75.4** The identification results of different test objects

Primary signal	Identification accuracy (%)
Lyf	91.6364
Hsl	92.0113
Lyf and Hsl	79.2365

**Table 75.5** The running time of different algorithms

Algorithm	Running time(s)
AR + Fisher	2.8
WT + SVM	3.6
This study	3.1

classification accuracy slightly declines; Table 75.4 indicate that individual difference exists in MI EEG classification, at the same time, the classification varies for different test sets; it shows in Table 75.5 that the algorithm put forward in this study has good real-time performance, which would provide the basis for real-time transmission and decoding of brain consciousness in online BCI system.

## 75.6 Conclusion

As for the nonlinear and non-stationary MI EEG, EMD based on the HHT method and Hilbot transform can effectively and fully demonstrate its effective information, and the classification effect of the nonlinear time varying SVM classifier is very good. The identification accuracy of two kinds of simple imagery motion is higher than multiple imagery motions. Single-trial identification of MI EEG based on HHT and SVM can improve the working performance and can meet the need of online BCI system for single-trial identification of EEG, having good identification accuracy and real-time performance, which would provide basis for the realization of online BCI control system based on MI EEG.

**Acknowledgments** The work for this paper was financially supported by the National Natural Science Foundation of China (NSFC, Grant No: 60841004, 60971110, and 61172152).

## References

1. Jin J, Zhang Y, Wang XG (2011) A novel combination of time phase and EEG frequency components for SSVEP-based BCI. *Neural Inf Proc* 7062:273–278
2. Ryu YS, Lee YB, Lee CG, Lee BW, Kim JK, Lee MH (2011) Comparative analysis of the optimal performance evaluation for motor imagery based EEG-brain computer interface. 5th Kuala Lumpur international conference on biomedical engineering, IFMBE Proceedings 2011, pp 488–491.
3. Duan LJ, Wang XB, Yang Z, Zhou HY, Wu CP, Zhang Q, Miao J (2011) An emotional face evoked EEG recognition method based on optimal EEG feature and electrodes selection. *Neural Inf Proc* 7062:296–305
4. He L, Gu ZH, Li YQ, Yu ZL (2010) Classifying motor imagery EEGs by iterative channel elimination according to compound weigh. *Artif Intell Comput Intell* 6320:71–78
5. Holper L, Wolf M (2011) Single-trial classification of motor imagery differing in task complexity: a functional near-infrared spectroscopy study. *J Neuron Eng Rehabil* 8(1):34
6. Aberg MCB, Wessberg J (2007) Evolutionary optimization of classifiers and features for single-trial EEG discrimination. *Biomed Eng Online* 6(1):32
7. Li Y, Koike Y (2011) A real-time BCI with a small number of channels based on CSP. *Neural Comput Appl* 20(8):1187–1192
8. Lee S, Lee S (2008) ICA-based spatio-temporal features for EEGs. *Neural Inf Proc* 4985:915–920
9. Huang NE, Shen Z, Long SR (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and no-stationary time series analysis. *Proc R Soc Lond* 454:903–995
10. Liao X, Yin Y, Li CY, Yao DZ (2006) Application of SVM framework for classification of single trial EEG. *Adv Neural Netw* 3973:548–553
11. Braecklein M, Pang L, Tchoudovski I, Kellermann W, Bolz A (2005) Comparison of two different methods for coherent averaging in online ECG analysis. *Comput Cardiol* 32:463–466
12. Yunfa F, Baolei X et al (2011) The study of single-trial identification of imagined movement speeds based on EEG. *Chin J Biomed Eng* 30(4):555–561(in Chinese)

# Chapter 76

## Robust Visual Tracking Using Incremental Sparse Representation

Song Pan and Huaping Liu

**Abstract** The sparse representation has achieved considerable success in visual tracking due to its simplicity and robustness. It requires each target candidate is sparsely represented in the space spanned by target templates and trivial templates. The sparsity is achieved by solving an  $l_1$ -regularized least squares problem. When the sparse representation is incorporated into the framework of particle filter, solving  $l_1$  minimization problem for each particle independently requires a large calculation time, making real-time implementation difficult. In this paper, we exploit the redundancy between particles and use the homotopy method to design an incremental likelihood function calculation approach, and therefore form an efficient and robust visual tracking algorithm. The proposed algorithm is tested on extensive video sequences and the experimental results are found to be highly competitive with other recent trackers.

**Keywords** Visual tracking · Sparse representation · Incremental approach ·  $l_1$  minimization

### 76.1 Introduction

Visual tracking has received significant attentions due to its crucial value in practical applications such as surveillance, video indexing, and so on. One popular approach to tackle the problem of visual tracking is to define it as sequentially

---

S. Pan

Department of Computer Software and Theory, School of Computer,  
Wuhan University, Wuhan, China  
e-mail: pans@ichaier.com

H. Liu (✉)

Department of Computer Science and Technology, Tsinghua University,  
Beijing FIT3-D02, China  
e-mail: hpliu@mail.tsinghua.edu.cn

estimating the state of a dynamic system using a sequence of noisy observations [1]. For this task, there has been immense attention on particle filters because, for any nonlinear or nonGaussian dynamic estimation problem, one can design recursive Bayesian filters by using the Monte Carlo sampling. It has been shown that for many tracking scenarios, constant subspace assumption is more reasonable than constant brightness or color assumptions [2, 3]. During the past decades many works had been developed to construct suitable likelihood function based on the subspace representation, but how to design a robust likelihood remains an open challenging problem.

Recently sparse signal reconstruction has gained considerable interests. Variations and extensions of sparse representation have been applied to many vision tasks, such as face recognition [4], image super-resolution, and video concept detection. In many fields, using sparsity as a prior leads to state-of-the-art results. Based on the basic assumption that good target candidate can be sparsely represented by both the target templates and the trivial templates, Ref. [5] incorporates the sparse representation into the tracking framework. This sparse optimization problem is solved as an  $l_1$  minimization problem with non-negative constraints. Then the candidate with the smallest target template projection error is chosen as the tracking result. In Ref. [6], a multipart subspace appearance model is developed for improving robustness and an overcomplete dictionary is learned off-line. These approaches utilize particle filter to realize the tracking and sparse representation is used to construct the likelihood function. Therefore an obvious problem is that for each particle an  $l_1$  optimization problem needs be solved and the computational burden is high. Although [7] proposes fast tracking algorithm with two stage sparse optimization, it essentially depends on an online selftraining classifier and easily suffers from drifting. In addition, the parameter setting in the proposed two-stage optimization [7] is also nontrivial. Very recently, Ref. [7] proposes an efficient L1-tracker with minimum error bound. In this approach, the minimum error bound is quickly calculated from a linear least squares equation, and serves as a guide for particle resampling in a particle filter framework. Without loss of precision during resampling, most insignificant samples are removed before solving the computationally expensive  $l_1$  minimization function. This approach enables us to speed up the original L1-tracker [5] for about 5 times when using 600 particles. However, when the object is occluded, the time-cost reduction becomes trivial.

In this paper, we exploit the redundancy between particles by employing the recently proposed homotopy approach to form a very efficient and robust visual tracking algorithm. The algorithm is tested on extensive image sequences and the experimental results confirm that the proposed approach is competitive with the state-of-the-art tracker but achieves obvious reduction in time costs.

The organization of this paper is as follows: Sect. 76.2 gives a brief review of the particle filter tracking approach. Section 76.3 gives the details of the proposed incremental sparse representation algorithm. Finally, we give extensive experimental comparison in Sect. 76.4.

## 76.2 MCMC-Based Tracking Approach

The task of tracking is to use the available measurement information to estimate the hidden state variables. Given the available observations  $Z_{1:k} = Z_1, Z_2, \dots, Z_{k-1}$  up to time instant  $k-1$ , the prediction stage utilizes the probabilistic system transition model  $p(X_k|X_{k-1})$  to predict the posterior at time instant  $k$  as  $p(X_k|Z_{1:k-1}) = \int p(X_k|X_{k-1})p(X_{k-1}|Z_{1:k-1})dx_{k-1}$ . At time instant  $k$ , the observation  $Z_k$  is available, the state can be updated as  $p(X_k|Z_{1:k}) = p(Z_k|X_k)p(X_k|Z_{1:k-1})/p(Z_k|Z_{1:k-1})$ , where  $p(Z_k|X_k)$  is described by the observation equation. Unfortunately, the integral in the recursive Bayesian estimation is analytically intractable and sampling approach should be used for approximation.

Markov Chain Monte Carlo (MCMC) proves to be an efficient way to achieve this goal [8]. In this case, a set of unweighted samples  $\{X_k^i\}_{i=1}^N$  generated by MCMC sampling is used for approximation of  $p(X_k|Z_{1:k})$ . The famous Metropolis-Hastings (MH) algorithm can be used to generate an unweighted sample set  $\{X_k^i\}_{i=1}^N$  with posterior distribution  $p(X_k|Z_{1:k})$ . In M-H algorithm, a proposed move is generated by the proposal distribution  $q(X'|X_k^i)$ . The move is accepted with an acceptance ratio  $a = \min\{1, \pi(x')q(x'|x_k^i)/\pi(x)q(x_k^i|x')\}$ . If accepted, the new particle  $X_k^{i+1}$  is set to be  $X'$ . Otherwise, the move  $x'$  is discarded and the new particle  $x_{i+1}$  remains to be the same as  $X_k^i$ . By this way, distribution of samples generated by MCMC will approximate desired distribution  $\pi(x)$ . In many cases, the proposal distribution  $q(X'|X_k^i)$  is symmetry, i.e., we have  $q(X'|X_k^i) = q(X_k^i|X')$  and therefore it reduces to be  $a = \min\{1, \frac{\pi(x')}{\pi(x)}\}$ . The desired distribution  $\pi(x)$  can be set as  $\pi(x) = p(Z_k|X) \sum_{i=1}^N p(X|X_{k-1}^i)$ , where  $p(Z_k|X)$  is the likelihood function and  $p(X|X_{k-1}^i)$  is the motion prior. The whole detailed sampling procedure can be found in [8]. If the iteration is finished, we can get a new sample set  $\{X_k^i\}_{i=1}^k$  and the estimated value of the state  $x_k$  can be approximated by the *Maximum A Posterior* (MAP):  $\hat{X}_k = \arg \max_i = 1, 2, \dots, N \pi(X_k^i)$ .

To enhance the tracking robustness, Ref. [5] proposed to use the sparse representation technology to construct the likelihood function. They construct the template library as  $B = [T I - I]$ , where  $T = [t_1, t_2, \dots, t_T]$ ,  $t_i$ , for  $i \in [1, T]$  is the template vector with appropriate dimension, and  $I$  is the unitary matrix with appropriate dimensions, which represents the trivial templates [5]. We can therefore reconstruct each candidate sample  $c_k^i$  with state  $X_k^i$  from the template library  $B$  using sparse representation. The likelihood is then set as  $p(Z_k|X_k^i) \propto \exp(-err_k^i)$ , where  $err_k^i = \|c_k^i - Ba_k^i\|_2$  is the optimized reconstruction error of  $c_k^i$ . In the above equation,  $a_k^i$  represents the sparse coefficient vector which solves the optimization problem  $\min a_k^i \|c_k^i - Ba_k^i\|_2 + \lambda \|a_k^i\|_1$ , where  $\|\cdot\|_1$  and  $\|\cdot\|_2$  denote the  $l_1$  and  $l_2$  norms respectively, and  $\lambda > 0$  is a regularization parameter. According to the implementation of Ref. [5],  $\lambda$  can be fixed as 0.01.

The whole tracking algorithm is listed in Algorithm 76.1 (The fifth line will be explained in the end of the next section). From this list we can see that the main bottleneck is the calculation of the sparse representation coefficient  $a_k^i$ . Since the number of the particles is  $N$ , we have to make the above calculation for  $N$  times and the computation burden is too large for real-time tracking. Fortunately, the MCMC sampling strategy is essentially a *sequential sampling* approach because the particle  $\tilde{X}_k^{i+1}$  is sampled from the neighborhood of  $X_k^i$ . We can therefore expect *the support of the sparse representation of  $X_k^i$  is similar to the support of the sparse representation of  $\tilde{X}_k^{i+1}$*  and we try to exploit this redundancy to reduce the calculation time cost. This is an important reason why we choose MCMC sampling, but not usual sequential importance sampling [5], under which all particles are generated in batch manner and the redundancy between particles is difficult to exploit. In the next section, we will give details on how to exploit the redundancy between particles.

### 76.3 Incremental Calculation of the Likelihood Functions

Denote  $\mathfrak{R}(X_k^i)$  and  $\mathfrak{R}(\tilde{X}_k^{i+1})$  to be the image regions associated with particle  $X_k^i$  and  $\tilde{X}_k^{i+1}$ , respectively. Then the problem can be reformulated as: Given an image region  $\mathfrak{R}(X_k^i)$ , and its sparse representation  $a_k^i$ , how to get the sparse representation  $a_k^{i+1}$  for image region  $\mathfrak{R}(\tilde{X}_k^{i+1})$ , which has large overlap region with  $\mathfrak{R}(X_k^i)$ . We denote the feature vectors for  $\mathfrak{R}(X_k^i)$  and  $\mathfrak{R}(\tilde{X}_k^{i+1})$  to be  $c_k^i$  and  $c_k^{i+1}$ , respectively, and then assume that  $a_k^i$  which solves the optimization problem

$$\min_a \|c_k^i - Ba\|_2^2 + \lambda \|a\|_1 \tag{76.1}$$

---

**Algorithm 1** MCMC-based particle filter tracking

---

Input:  $\tilde{X}_{K-1}$

Output:  $\tilde{X}_K$

---

- 1: Initialization: Draw the first particle  $X_k^{(1)}$  from the distribution  $q(\tilde{X}_K | \tilde{X}_{K-1})$ ; Extract the feature vector  $c_k^{(1)}$ ; Extract the sparse representation  $a_k^{(1)}$  for  $c_k^{(1)}$  and get the weight.
  - 2: for  $l = 1, 2, \dots, N - 1$  do
  - 3: Draw predicted particles from the proposal distribution  $\tilde{x}_k^{i+1} \sim q(x^l | x_k^i)$
  - 4: Extract the feature vector  $c_k^{i+1}$  for the particle  $\tilde{x}_k^{i+1}$ .
  - 5: Extract the sparse representation  $a_k^{i+1}$  for  $c_k^{i+1}$  and get the weight.
- 

(continued)



(continued)

---

**Algorithm 1** MCMC-based particle filter tracking

---

6: Calculate the acceptance ratio  $a$ , where  $x'$  is replaced by  $\tilde{x}_k^{i+1}$ . Then we accept this predicted particle according to the probability  $a$ , i.e., let  $x_k^{i+1} = \tilde{x}_k^{i+1}$ . If this predicted particle is rejected, then we let  $x_k^{i+1} = x_k^i$ .

7: Update the template library  $B$  according to [5].

8: end for

9: Obtain the MAP estimation  $\hat{X}_K$ .

---

is the sparse representation of the sample  $c_k^i$ . Similarly, we denote  $a_k^{i+1}$  to be the sparse representation of the sample  $c_k^{i+1}$ . It can be solved by directly solving the following optimization problem

$$\min_a \|c_k^{i+1} - Ba\|_2^2 + \lambda \|a\|_1 \quad (76.2)$$

However, the optimization in (2) is slow and if the number of the particle  $N$  is large, the time-cost of the tracking will be huge. Motivated by the fact that the region  $\mathfrak{R}(\tilde{X}_k^{i+1})$  is close to  $\mathfrak{R}(X_k^i)$ , we can expect that the extracted sample  $c_k^{i+1}$  should be similar to  $c_k^i$ . So, a natural intuition is proposed that there will be little difference between  $a_k^{i+1}$  and  $a_k^i$ . Therefore we can solve  $a_k^{i+1}$  based on  $a_k^i$ . This can be easily realized by using homotopy continuation, which is recently proposed in [1]. This approach is very much like recursive least squares: In each step, only a low-rank update and a small number of matrix-vector multiplications are involved. It is very effective when the support of the solution does not change too much from particle to particle. However, it should be noticed that Ref. [1] does not deal with visual tracking problem.

Assume that calculating time using  $l_1 - l_s$  and the above-mentioned homotopy approach for one particle is  $T_{l_1}$  and  $T_h$ , respectively, then the time costs is  $NT_{l_1}$  for [5] and  $T_{l_1} + (N + 1)T_h$  for our approach, where  $N$  is the number of particle. In our implementation,  $T_h$  is only about  $1/9$  of  $T_{l_1}$  and therefore the computational advantage is obvious. More detailed time complexity analysis can be found in [1].

## 76.4 Experimental Validation

When implementing our approach, we should set some parameters. In all of the experiments, we set  $\lambda = 0.01$ . The state of each particle filter is defined as  $X_k = [x_k \ y_k \ s_k]^T$ , where  $x_k$  and  $y_k$  indicate the locations of the object;  $s_k$  is the corresponding scale. The proposal distribution is designed as a simple 3-dimensional zero-mean Gaussian distribution. The variance vector is set as  $[1 \ 1 \ 0.01]^T$ , of which the first two component correspond to the location and the third one

corresponds to the scale. In the experimental validation, all algorithms are the third one corresponds to the scale. In addition, all algorithm parameters are fixed for all the experiments.

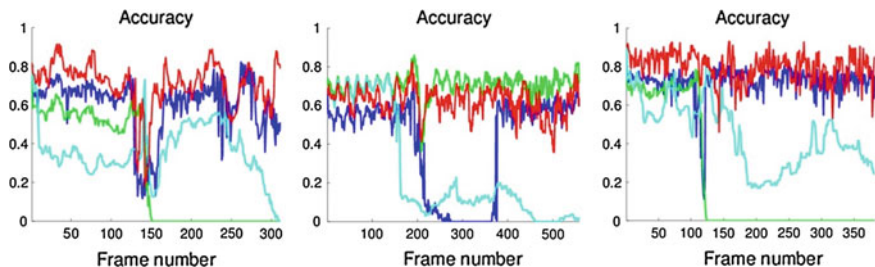
To quantitatively evaluate the absolute performance, we use CAVIAR dataset (downloaded from <http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/>) for evaluation since it provides complete ground-truth data. Besides L1-tracker, we also compare our approach to IVT [2, 3] and ensemble tracking (ET) [9], using code provided by the respective authors. Note that ET is based on mean-shift search but IVT, L1-tracker and our approach are based on particle filter and the number of particle is set to 50.

The experiments are conducted on 15 representative sequences of CAVIAR dataset. Due to the motion intrinsic of the pedestrian to be tracked, we modify the original code of IVT and L1-tracker by setting the parameters such that only location and scale are tracked, rather than a full set of affine parameters.

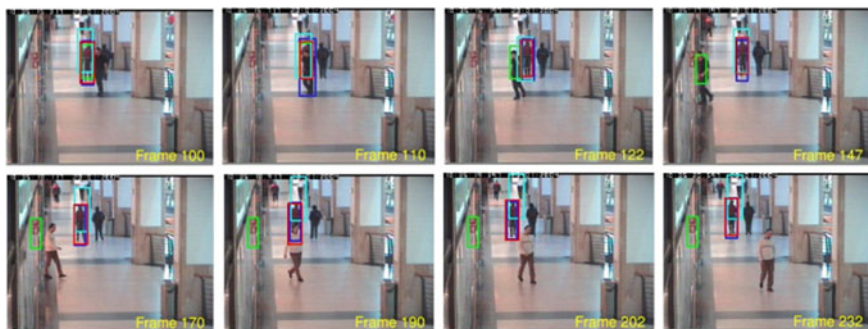
The tracking accuracy for each frame is evaluated by the overlap ratio, which is defined as the ratio between the area of intersection  $GT \cap TR$  and the area of union  $GT \cup TR$ , where GT and TR are the ground truth box and the tracking result box, respectively. The mean tracking accuracy over the whole sequence are given in Table 76.1, where the column ID represents the object ID which is defined according to CAVIAR ground-truth file, Frm# is the number of the frames. In Table 76.1 the *red* number indicates the best tracking accuracy and the *blue* number indicates the second tracking accuracy. From these results we see that the sparse representation approaches (L1-tracker and the proposed approach) achieve rather good performance. In addition, the proposed approach is better than L1-tracker except sequence *ShopAssistant2cor*. The partial reason is that the

**Table 76.1** Accuracy evaluation

Sequence information		Mean tracking accuracy			
Sequence name	ID Frm#	L1-tracker	IVT	ET	OURS
<i>EnterExitCrossingPaths1cor</i>	0 382	0.7203	0.2119	0.4457	0.7974
<i>EnterExitCrossingPaths1cor</i>	0 586	0.7003	0.3404	0.326	0.7515
<i>EnterExitCrossingPaths1cor</i>	2 234	0.7112	0.1878	0.5353	0.7481
<i>OneLeaveShopReenter2cor</i>	0 560	0.4065	0.7081	0.256	0.6228
<i>OneShopOneWait1cor</i>	3 273	0.3817	0.5691	0.4104	0.5381
<i>OneShopOneWait2cor</i>	7 1266	0.3802	0.048	0.4307	0.4737
<i>OneStopEnter1cor</i>	1 585	0.6098	0.2262	0.357	0.6728
<i>OneStopEnter1cor</i>	1 448	0.7153	0.3847	0.5002	0.7932
<i>OneStopNoEnter2cor</i>	0 855	0.6231	0.4588	0.3997	0.7258
<i>ShopAssistant2cor</i>	18 772	0.629	0.173	0.1221	0.446
<i>ThreePastShop1co</i>	0 325	0.7181	0.5123	0.0592	0.7401
<i>ThreePastShop1co</i>	2 310	0.5983	0.2542	0.3594	0.7072
<i>TwoLeaveShop1cor</i>	3 592	0.1502	0.4927	0.1165	0.6522
<i>TwoLeaveShop2cor</i>	2 252	0.1425	0.3963	0.3797	0.5249
<i>WalkByShop1cor</i>	5 352	0.4172	0.4454	0.4633	0.5613



**Fig. 76.1** Tracking accuracy comparison: from left to right: *EnterExitCrossingPaths1cor*, *OneLeaveShopReenter2cor*, *OneLeaveShopReenter2cor*. Blue L1-tracker; Green IVT approach; Cyan ET approach; Red Proposed approach



**Fig. 76.2** Tracking results of *EnterExitCrossingPaths1cor*. Blue L1-tracker; Green IVT approach; Cyan ET approach; Red Proposed approach

proposed approach intrinsically encourages that similar samples share similar sparse representations and therefore improve the robustness.

As to the tracking time, we can see that IVT is the fast approach, but its accuracy is not satisfactory for most sequences. Our approach's time cost is similar to ET approach while L1-tracker is the slowest approach. In fact, when the number of particle is set to 50, the time costs of L1-tracker is about 4–5 times than ours. Even using a MATLAB implementation, the proposed method can process near two frames per second.

## 76.5 Conclusion

The main contribution of this paper is to propose a robust MCMC-based visual tracking algorithm which incorporates the incremental sparse representation calculation. Extensive experimental results validate the efficiency of the proposed approach. The proposed algorithm is tested on extensive video sequences and

the experimental results are found to be highly competitive with other recent trackers. In the future we will extend this work to multiple-object tracking (Figs. 76.1, 76.2).

**Acknowledgments** This work is jointly supported by the National Key Project for Basic Research of China (2013CB329403), the National Natural Science Foundation of China (Grants No:61075027, 91120011, 61210013) and the Tsinghua Selfinnovation Project (Grant No: 20111081111).

## References

1. Perez P, Hue C, Vermaak J, Gangnet M (2002) Color-based probabilistic tracking. In: Proceedings of European conference on computer vision, pp 661–675
2. Ross DA, Lim J, Lin R, Yang M (2008) Incremental learning for robust visual tracking. *Int J Comput Vis* 77(1):125–141
3. Asif MS, Romberg J (2010) Dynamic updating for  $l_1$  minimization. *IEEE J Sel Top Signal Process* 4(2):421–434
4. Wright J, Yang A, Ganesh A, Shastry S, Ma Y (2009) Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell* 31(2):210–227
5. Mei X, Ling H (2009) Robust visual tracking using  $l_1$  minimization. In: Proceedings of international conference on computer vision, pp 1–8
6. Zhang J, Cai W, Tian Y, Yang Y (2009) Visual tracking via sparse representation based linear subspace model. In: Proceedings of international conference on computer and information technology, pp 166–171
7. Liu B, Yang L, Huang J, Meer P, Gong L, Kulikowski C (2010) Robust and fast collaborative tracking with two stage sparse optimization. In: Proceedings of European conference on computer vision, pp 624–637
8. Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov Chain Monte Carlo in practice. Chapman and Hall, London
9. Avidan S (2007) Ensemble tracking. *IEEE Trans Pattern Anal Mach Intell* 29(2):261–271

# Chapter 77

## A Study on Design of Early Warning Decision Support System of Desertification

Zhengwei Li, Jing Du, Xianyong Meng, Chen Sun and Yongqiang Liu

**Abstract** At present, the study on the slowly varying monitoring and early warning system is not perfect at home and abroad. In this paper, desertification database, knowledge base and model base are designed in details and the future evolution of the desert is predicted by the desertification early warning system. The results show that correct rate of simulated distribution of desertification reaches over 90 % in the study area. And ‘Auto Set Sand instrument—Monitoring Platform’ can conduct sand data analysis at anytime. In addition, the WebGIS based system provides necessary decision support for the government.

**Keywords** Desertification early warning · Database · Model-library · WebGIS · Decision support system

### 77.1 Introduction

According to land desertification trends and real-time performance, combined Computer software technology, GIS technology with communication technology, this paper provides real-time and reliable continuous data for the desert research.

---

X. Meng · C. Sun · Y. Liu (✉)  
School of Resources and Environment Science, Xinjiang University, Urumqi 830046, China  
e-mail: Lyq.xj@tom.com

X. Meng · C. Sun · Y. Liu  
Key Laboratory of Oasis Ecology Ministry of Education, Xinjiang University, Urumqi 830046, China

Z. Li  
Environment Science Technologies Co.,Ltd. of Yunnan Luhong, Yunnan 650228, China

J. Du  
School of Computer Science and Technology, Xinjiang normal university, Urumqi 830054, China

The system includes three parts: data collection terminal (Collects Sand Instrument), monitoring center software system and web publishing platform. These three parts can complete the data collection and management, data processing and model analysis, monitoring information and release the results of early warning. Ground monitoring mainly refers to using automatically collecting sand instrument timing to obtain the weight of sand, wind speed, wind direction, temperature four elements which closely related to the sand and return data through the communication network to the data center regularly. Data collection terminal provides data sources for the entire system. Monitoring center software system is the core of desertification monitoring and warning system and has desertification database and modelbase and knowledgebase of Xinjiang which provide database and decision support for system. Web publishing side is a publishing platform of desertification monitoring and warning system which releases site real-time monitoring data, desertification forecast warning in study area, monitoring data and early warning analysis etc.

## **77.2 The Design of Desertification Monitoring System**

### ***77.2.1 System Software Architecture***

Considering system users and performance, the system uses C/S and B/S architecture combined and can be divided into four layers: application service layer, business layer, data layer and infrastructure layer. Application service layer is responsible for visual display and interacting data with user, users achieve all the necessary actions through the monitoring center client or WEB browser sending a request to the network server. The business layer is responsible for response and processing user requests to finish lots of data analysis which achieve by the Super Map IS.NET and IIS, etc. Data layer is responsible organization and management for data and accept the request of which server operates on the database. This system architecture creatively adds the data acquisition subsystem (automatically collecting sand instrument), which sends data to the monitoring center and accepts it's control (Fig. 77.1).

### ***77.2.2 The Design of Overall System Features Modular***

The system relies on the support by monitoring equipment and computer software and hardware, establishes a monitoring center platform based on C/S structure and provides functions for professionals including receiving and managing the field real-time monitoring data, database management, data processing and analysis, model management, computer applications of model and analysis of monitoring results, etc. It establishes an information publishing platform based on B/S

Application Layer	Data monitoring system	Monitoring center system	Information release system
Middle layer	GIS /SDB+Engine	Communication interface	Data interface
Data layer	Desert Database	Knowledge database	Model library
Infrastructure layer	Operating System	Supporting software	Database
	Communication facilities	Data acquisition equipment	Servers and storage devices

Fig. 77.1 The structure of software system

structure, provides users with features including monitoring data published, data inquiry, early warning published and map analysis, etc. through background management system (Fig. 77.2).

**77.2.2.1 Data Monitoring Subsystem**

Data Monitoring Subsystem receives data and stores data in Desertification Database and conducts other functions including service control, system parameter settings, site registration and data analysis, etc.

**77.2.2.2 Monitoring Center Subsystem**

Monitoring Center Subsystem is the core of the desertification monitoring and warning system, center staff can manage on background data and data analysis by monitoring center system.

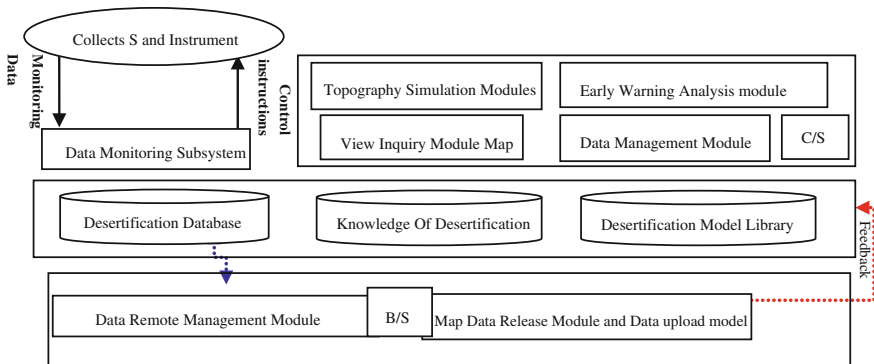


Fig. 77.2 The structure of overall system features modular

Data management module: include management of database, knowledge base and model base.

Map Browse and Inquiry Module: displays the map images and vectors of study area through the map window and combines graphic data and attribute data in order to enable users to conduct data inquiry and data analysis conveniently.

Desertification forecast and Analysis module: predicts the trend of desertification of the study area based on prediction model.

Topography Simulation analysis modules: simulate terrain of study area and provide a variety of three-dimensional analysis functions.

### **77.2.2.3 Information Distribution System**

This system refers to the desertification monitoring and early warning sites, mainly used to publish with desertification-related news, the results of early warning and real-time monitoring data and provide data query including spatial data, monitoring data and basic data meanwhile allow the system administrators to manage the database remotely.

### ***77.2.3 Early-Warning Database***

Xinjiang Desertification Database not only includes spatial data associated with desertification, but also real-time monitoring data, historical statistical data, etc. The entire database is divided into three parts: Metabase, Spatial Database and Attribute Database.

1. Metabase: there are variety of data types in Xinjiang Desertification Database, system establishing Metadata table of Spatial data and Metadata table of Attribute data to Presentation contents of data, quality, data formats, data owners, etc. [1].
2. Spatial Database: stores the vector, raster, digital elevation with desertification-related and image data scanned maps. Spatial database is also the basis of decision support systems [2].
3. Properties Database: the database not only stores real-time monitoring data of Collecting Sand Instrument but also stores multiple database tables which relate to Xinjiang's climate data, surface factors, social and cultural data and through keyword to be relevant.

Attribute data mentioned in this study is relative to the spatial data, basic property data and real-time monitoring data. Spatial data and the basis of attribute data can be associated through keywords, Such as the Entity ID, NAME, etc. For the association between Spatial data and Monitoring data (for generate real-time monitoring Figure), System through the unambiguous ID code of spatial objects



(Such as the site ID) association with attribute data fields in the table. There is a certain relationship between the monitoring data table and Spatial data table and Monitoring data table.

### ***77.2.4 The Construction of Early Warning Models of Desertification and Model Libraries***

Model library is a core part of the Decision Support System, It's used to drive Decision Support System for decision-making. Model library refers to a collection of stored in the computer according to a certain structure [3], models can extract the laws of reality, and establish the relationship between the rules to simulate reality of the future. This paper sets early warning models of desertification in the middle lower reaches of Tarim area as an example of model structure [4]:

$$SM = \eta \cdot \sum_{i=1}^n (Q_i \cdot W_{ci}) \quad (77.1)$$

$Q_i$  (factor intensity),  $W_{ci}$  (factor weight coefficient),  $SM$  (desertification index) and  $\eta$  (adjustment coefficient). When warning is achieved,  $Q_i$  uses raster  $i$  factors figure form,  $W_{ci}$  With  $\eta$  is a group of parameters. The formation and development of land desertification is the result of many factors, because the causes of desertification in different parts are different, so a fixed models is difficulty to apply in all regions. Therefore, various research areas should establish the corresponding models to improve the simulation accuracy. Early Warning Models of Desertification can be split into three parts: relationship, parameters and variables. Relationship is the relationship between the expression. Parameter is used to adjust the every variable indicator (Relational database has been conducted data storage), In order to adapt to different situations., variables are input interface of model, different regions may correspond to different models, The same area in different years corresponds to a different parameter sets.

The model storing, relations and parameter sets are stored separately in this study. Relationship of models is stored in the program module. Parameter sets are adopted parameter list and stored in relational databases which achieve an integrated storage between spatial Graphic data and attribute data [5].

#### **77.2.4.1 Relational Storage of Models**

In order to facilitate the program calls and expression of models, variables and parameters are used in the form of an array. The variable is obtained from the selection of list box, so for the list box control using array control forms and parameters can be read from the model library. The example of early warning

models of desertification in the middle lower reaches of Tarim area shows a relational storage of models program modules.

#### **77.2.4.2 The Storage of Model Parameter**

The parameter sets Early Warning Model of Desertification are classified by year, since index factors on the impact of desertification will gradually change, each year there will be a corresponding parameter sets. Although very little amount of relationship between the models, each relation corresponds to a very large set of parameters. Therefore, the organization and management of parameter sets are the key to early warning model library of desertification.

#### ***77.2.5 The Construction of Knowledge Repository Early Warning of Desertification***

Repository is knowledge base of Early Warning System to monitor desertification in Xinjiang, providing knowledge, experience and decision support for land desertification monitoring and early warning. The repository use three types to store: meta-knowledge table, knowledge of examples and rule knowledge table. For instance, early warning indicators of desertification and the summary of desertification, disaster management methods and policies is established based on Meng et al. (2013) [4, 6].

### **77.3 The Implementation the Systems of Desertification**

#### ***77.3.1 Data Monitoring Subsystem***

Data monitoring sub-system is the system platform which can receive field monitoring data and set associated with communication parameter, control data acquisition instrument remotely. Data transmission components of field data collection instrument use DTU Products of Shenzhen-Hong Dian Company. Development kits of the DSC matched to DTU are the basis for system development. Functions Include communications parameter settings (service control, system settings, DTU management, etc.). In addition, a number of functions are added with the actual needs of the project (Data interpretation and storage, remote control of Collecting Sand Instrument, data storage set, and the site registration, etc). And a data interpretation and storage mechanism is established to convert the communication data.

### ***77.3.2 The Design and Implementation of Forecast and Results Analysis Module of Desertification***

The early warning analysis module of desertification is the core of desertification monitoring and warning system. The core functions of the module include: desertification forecast, results analysis and desertification early warning function.

Desertification prediction and results analysis: forecast of desertification is the basis for early warning and implemented through prediction model of desertification. Select determined model parameters first, and then determine the status of data variables, including model parameter settings and model computing settings. Model computing settings: open the raster image, select the appropriate variables of operator for predict desertification, and set prediction data names and storage location.

### ***77.3.3 Desertification Early Warning***

Early warning of desertification extract the area which land situation changes after a comparative analysis between desertification predicted and desertification status data. There are three types of output for the user to choose. System extracts the area of desertification early warning based on user inputted the critical value.

Three types of result:

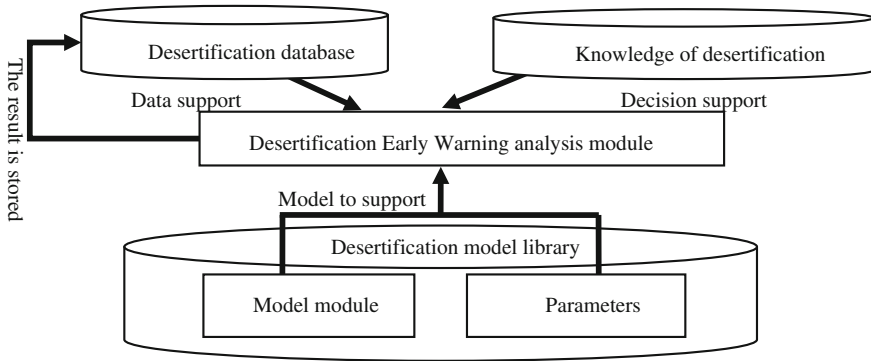
Condition improved area of land desertification. The type of data refers to the region which will reverse the development of land desertification in the future, Land conditions have been improved and desertification was reduced. Naming rules: [forecast year] + melioration + [self named].

Sustainable development area of land desertification. The type of data refers to the region which will increase the degree of desertification. Naming rules: [forecast year] + deterioration + [self named].

Early warning area of desertification. The type of data refers to extracting the region where desertification surpasses the warning value from sustainable development area of land desertification. Naming rules: [forecast year] + alarm + [self named].

### ***77.3.4 Web Publishing Subsystem***

Network publishing subsystem is based on the B/S structure, with Supermap IS.NET as GIS development platform and development environment in Visual Studio 2008. The subsystem has user registration, message boards, vector map browse, etc. Users can browse and query the associated vector map, image data, early warning map and related attribute data. System uses WebGIS technology to



**Fig. 77.3** Early warning analysis module structure of desertification

establish platform for data query and browsing and query monitoring and early warning data through a web page (Fig. 77.3).

**Acknowledgments** To express my gratitude for the research projects were sponsored from the R&D Special Fund for Public Welfare Industry (Meteorology) (GYHY201306066).

## References

1. Ma L, Zhang X (2010) Principles and methods of WebGIS. Science Press, Beijing 3–5, 215–219 (in Chinese)
2. Qi P, Liu W (2009) Research and realization of WebGIS-spatial data. *Sci Technol* 13:446–447 (in Chinese)
3. Chen D (2005) The study on model library in decision support system. Harbin Engineering University, Harbin (in Chinese)
4. Meng X y, Liu Z et al. (2013). Sand desertification early-warning model based on a 30 M\*30 M gridscale. *J Desert Res* 33(1):24–32 (in Chinese)
5. Wu X (2009) Spatial database. Science Press 30 May 2009 (in Chinese)
6. Liu D, Zheng J, Liu Z et al (2009) Spatial-temporal simulation and prediction of sandy desertification evolution in typical area of Xinjiang. In: Proceedings of the China Ireland information technologies conference, (Maynooth, Ireland, 2009) 19–21st, pp 90–96

# Chapter 78

## Fast Fusion Method of TT&C Data with Multi-Routing Transmission Model

Bin Tian, Yue Yang, Yanhui Pan and Shengjun Luo

**Abstract** In view of the problem of valid data identification brought by the multi-routing transmission model in the new generation of TT&C IP network, the characteristic of TT&C IP network is analyzed. According to the requirements of transmitting TT&C data in real time and with high reliability, the short delay priority principle is taken, the comparison operator of TT&C data is defined, and the algorithm of fast fusion of multi-routing data suitable for the receiver node is given and tested. The results show that the algorithm has strong points of short delay and high real-time performance.

**Keywords** TT&C · Data fusion · Multi-routing · Data flow

### 78.1 Introduction

The aerospace TT&C data is transmitted by specified channel traditionally. With the rapid development of aerospace industry, the interval between aerospace missions becomes shorter and shorter, which induces a surge in amount of data exchange through aerospace TT&C network. Since the traditional data

---

B. Tian

School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

B. Tian · Y. Pan (✉) · S. Luo

Xi'an Satellite Control Center, Xi'an 710043, China

e-mail: yanhuipan@126.com

B. Tian · Y. Pan · S. Luo

State Key Laboratory of Astronautic Dynamics, Xi'an 710043, China

Y. Yang

College of Software, Sichuan University, Chengdu 610065, China

transmission mode is difficult to satisfy the requirement of increasing aerospace missions, the aerospace IP TT&C network is proposed, which adopts packet switching technology. But just providing best effort service, the IP TT&C network is of some limitation in the aspect of reliability for the demands of aerospace TT&C data with high-precision [1]. In order to ensure the high reliability of the TT&C data transmission, usually the multi-routing mode is used in the new generation of IP TT&C network, which has advantages for redundancy and complementarity. But as a result of different network delays due to different routings, there would be delay difference, disorder or even data during the process when the same data transmitted to the receiver node. Furthermore, data redundancy may result in operation repletion of the same TT&C control instruction, leading to non-expected control or error results. Therefore, it is a key issue, which should be addressed by the receiver, to realize fast fusion of the multi-channel data generated by the source node, and finally to acquire the valid data.

## 78.2 Discuss

Multiple routing of TT&C network refers to multiple independent transmission channel in separate WAN. It means that the two WAN nodes are connected by multiple independent routing. And both node ends of the independent routing belong to different user's IP address space. Hence, there are plurality of paths between the source node and the receiver node for TT&C data transmission, and each path is usually referred as a route, as illustrated in Fig. 78.1. Wherein,  $\text{Packets}_s = \{P_1, P_2, \dots, P_i, \dots, 1 < i < \infty\}$  is a symbol that identifies the sequence of information frames, generated by the source node.  $\alpha \in [1, n]$ , is the route identifier. And  $n$  is the total number of routes.  $\text{Packets}_{s\alpha}$  identifies the sequence of information frame assigned to of the route  $\alpha$ .  $\text{Packets}_d$  identifies the sequence of information frame received by the receiver node. For certain aerospace missions, in multi-routing transmission mode, partial or all data are transmitted simultaneously from the source node to the receiver node by multiple paths. In other words, several copies of the same frames, generated by the source node, would be transmitted to the receiver node by diverse routes. At present, the dual-route transmission mode is used frequently [2]. Usually, the fusion of multi-routing data could be attributed to comparison of dual-routing data according to dichotomy. The dual-routing mode usually adopts the following to kinds of transmission model:

- (1) The homogeneous mode, the source node transmits the same information frame through each of the communication route simultaneously. It accords with the equation:  $\text{Packets}_{s1} = \text{Packets}_{s2} = \text{Packets}_s$ .
- (2) The heterogeneous mode, one route between the source node and the receiver node transfers the data frame, and the other one send frame of link monitoring

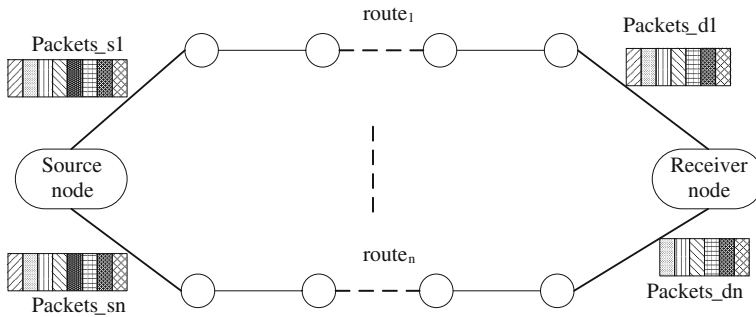


Fig. 78.1 The transmission model using multi-routing for TT&C data

information. It accords with the equation:  $Packets_{s1} \neq Packets_{s2}$ , and there are  $Packets_{s1} \cup Packets_{s2} = Packets_s$ .

For heterogeneous mode, frames are identical and they are transmitted to the receiver node through identical path. However, the homogeneous mode would cause the same problem. The same frames, produced by the source node, when reach the receiver node, they will not be identical taking account of the network character of frames. When TT&C data are transmitted from the source node to the receiver node, the network delay is variable. It always changes with the change of the network operation situation [3]. That is the main factors which should be considered when discussing the fusion method of TT&C data. The features of the IP packet switching network delay have been studied by numerous researchers [4–6] Usually, for different scales, the changes of the IP packet network delays are of certain rules and the delay jitter is bounded.

To solve the problem, a mechanism to fuse multi-channel data and obtain useful information frame should be established in the receiver node. Traditionally, the method of data fusion is mainly applied in the multi-sensor field. But recently, it has been gradually extended to other application areas [7, 8]. The multi-routing data fusion of TT&C network mainly is to analysis and processing time-series routing data obtained by the receiver node, and to gain the useful information frame sequence generated by the source node.

### 78.3 Fusion

In order to meet the multi-route transmission demands in aerospace TT&C field, a fast fusion method of TT&C data with multi-routing transmission model is proposed. The fusion principle and algorithm will be given in the following sections.

### 78.3.1 Fusion Theory

In multi-routing transmission mode, the discordance of data in the receiver node is mainly caused by the different features of various routing and the randomness of network transmission delay, in which the data lost can be considered with infinite delay. Therefore, the delay and the consistency of the multi-routing data are the main factors that should be considered.

#### 78.3.1.1 Delay Constraints

There would be various complex situations in multi-routing transmission mode because of different network delays of different paths. It shows that the LAN network delay is less than 10 ms in TT&C network and large scale WAN delay is up to several hundred milliseconds, according to measurement experiments based on the ICMP protocol. Let  $TE(P_i)$  identify the delay of frame  $P_i$ , which is the difference value between the frame received time and sent time. The principle of giving higher priority to shorter delay frame is adopted to ensure data are processed in real time. According to the transmission characteristics and application requirements of TT&C network, the value of TE should be in the range:  $1s \leq TE \leq 2s$ . Namely, if the value of TE is out the range, the corresponding data is useless.

#### 78.3.1.2 Comparison Operator

The purpose of multi-routing data fusion is to eliminate the redundant data, and select the useful data from multi-channels in the receiver node. Based on the above delay constraints principles, of all the frames with the same information reached in the receiver node, all the rest frames are redundant, excluding the frame with the shortest delay. Whether the frames are identical or not is judged by the comparison operator. The binary comparison operator  $C(P_i, P_j, G)$  is defined for frame information. Wherein,  $1 < j < \infty$ , and G is for comparison level. According to the transmission format of TT&C data, G is defined into three levels, listed in the follows:

The first level: partial data fields of  $P_i$  and  $P_j$  are compared, such as source identifier, destination identifier, task code, data type, data length, and so on. The operator value is true when the selected partial data fields are identical, otherwise the operator value is false.

The second level: the header fields of  $P_i$  and  $P_j$  are compared, such as time stamp, frame identifier, frame serial number, and so on. The operator value is true when the values of the selected header fields are identical, otherwise the operator value is false.



The third level, the entire data fields of  $P_i$  and  $P_j$  are compared. The operator value is true when the values of the entire fields are identical, otherwise the operator value is false.

All of the above, the first level is the simplest comparison, which can be realized directly by the comparison of primitive data types, such as string, integer and so on. But the accuracy level is low. The second level is limited in application since the frame structures of different TT&C objects are various. The third level is the most rigorous comparison and also the best one in practical application, which could be accomplished directly by comparison between blocks of memories. Because of mature and efficient algorithms, it is not difficult for computers to compare two memories now and the time-consuming of comparison between memories is not more than several times of comparison between integers.

### 78.3.1.3 Buffer Depth

For fusion procedures, TT&C data frames could be considered as streaming data. To compare information frame, a certain amount of historical data should be cached. For multi-routing fusion algorithm with high real-time requirements, the buffer depth is a critical factor affecting the processing speed. Usually, queue is used to cache the information frame. Let symbol L represent the buffer length, then:

$$\lceil TE_{\min} \times f_{source} \rceil \leq L/n \leq \lceil TE_{\max} \times f_{source} \rceil \quad (78.1)$$

Above symbol  $f_{source}$  presents the number of valid information frames generated by the source node per second. The unit of L is the length of one message frame. For example, taking  $f_{source} = 40$ ,  $n = 2$ , the value of L is an integer within the range of [80, 160].

### 78.3.2 Algorithm Design

Based on the above data fusion principles, a fast fusion algorithm is designed for TT&C data. The main procedure is as follows:

- (1) Initialization. Set the initial value of the maximum delay TE that the system can accept. Set periods TP's initial value to scan data buffer queue. And there is a condition:  $0.25s \leq TP \leq 1s$ . Start timed interrupt event Timer(TP). Construct frame buffer queue named buffer and initialize it. Set the initial value of L, which is the length of the information frame buffer queue. Set default value of the comparison level G.
- (2) Let  $i = 1$ , and  $i$  is the identifier of the received information frame.
- (3) Receive information frame  $P_i$  from any communication routing.

- (4) Check whether the delay of the information frame  $P_i$ .  $TE(P_i)$  is greater than  $TE_{max}$ ? If yes, discard the information frame  $P_i$ , and goes to step (8). Otherwise, proceed to step (5).
- (5) Check whether the information frame  $P_i$  is a redundancy or not.  

```
bool bExist = false;
for(j = 0; j < buffer.length; j++){
  if(C(buffer[j],  $P_i$ , G){
    bExist = true;
  }
}
```

 If the variable  $bExist$  is true, goes to step 6. Otherwise, proceed to step (7).
- (6) Process the redundant information frame. Discard the current information frame  $P_i$ .
- (7) Process useful information frame. Output the current information frame  $P_i$ , and insert the current frame  $P_i$  into the frame queue buffer.
- (8)  $i = i + 1$ , and goes to step (3).

In the process from step 1 to step 8, there is a specific process performed in cycle: when the timer interrupt event  $Timer(TP)$  generate an interrupt signals, check all the message frames buffered in queue buffer. If which the difference between the current time and the information frame sending time is larger than  $TE_{max}$ , delete it.

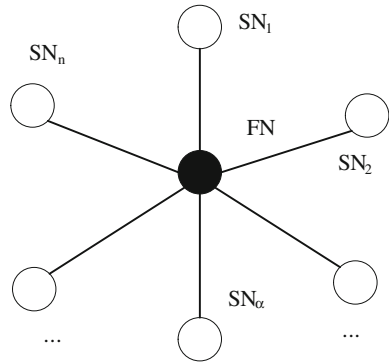
## 78.4 Analysis and Verification

### 78.4.1 Algorithm Effectiveness

#### 78.4.1.1 Simulation

According to the multi-route transmission mode shown in Fig. 78.1, the data fusion algorithms need to process data simultaneously from several routing direction. So a network structure shown in Fig. 78.2 is used to simulate the above algorithm. Let  $n$  SN (Source Node) send  $f_{source}$  same data to FN (Fusion Node) per second. But the sending patterns of information frames are various, simulating different delay features caused by different path. Fusion algorithm runs in the node FN. In the experiment, taking  $N = 4$ ,  $f_{source} = 8$ ,  $tp = 0.5s$ ,  $TE = 1s$ ,  $L = 40$ , and  $G = 3$ . The symbol  $Packets\_s$  represents information frames generated per second, tagged by number 1–8. Wherein, let  $SN_1$  and  $SN_2$  node send data normally, but part of the data are delayed.  $SN_3$  doesn't transmit data with certain probability.  $SN_4$  transmit data only at special moments. Thereby, the above experiment could simulate the coexistence of homogeneous and heterogeneous pattern. The specific settings are shown in Table 78.1.

**Fig. 78.2** The simulation network structure



**Table 78.1** The data transmission model of SN node

Node identifier	Data model
SN <sub>1</sub>	(1) {2} {3} (4) (5) {6} (7) {8}
SN <sub>2</sub>	{1} (2) (3) (4) {5} {6} (7) {8}
SN <sub>3</sub>	(1) (2) <3> (4) (5) {6} <7> (8)
SN <sub>4</sub>	<1> <2> (3) <4> <5> (6) <7> <8>

Remarks:  
 () means that send data normally.  
 {} means that send data with 10 ms delay.  
 <> means that not send data at the transmission time.

The output of FN node is 1-SN<sub>1</sub> → 2-SN<sub>2</sub> → 3-SN<sub>2</sub> → 4-SN<sub>1</sub> → 5-SN<sub>1</sub> → 6-SN<sub>4</sub> → 7-SN<sub>1</sub> → 8-SN<sub>3</sub>. It satisfies the condition of data fusion: Packets<sub>s</sub> = Packets<sub>d</sub>. Besides, the max length of information frame buffer queue is only seven during the running time.

**78.4.1.2 Practical Application**

Since the beginning of 2010, the algorithm runs in the foreign exchange core node of Xi'an Satellite Control Center. After numbers of missions, the algorithm is proved that it runs in good condition, achieving the purpose of fusing multi-routing TT&C data quickly in real-time.

**78.4.2 Algorithm Performance**

The buffer queue in the algorithm only storage the output useful data and the algorithm clears the overtime data periodically, so the calculation time cost is a constant and do not increase with the growth of data volume. In addition, the input

data are compared only with the data in buffer queue and they will not stay so that they are not compared with other data. Therefore, the algorithm is of the advantages of short delay, high real-time performance, and giving the fusion results in the fastest speed among the multi-routings at any time.

## 78.5 Summary

This paper researches and analyzes the multi-routing transmission model of TT&C data, and proposes a fast fusion method used for obtaining the useful information frames generated by the source node. Meanwhile, the implementation steps of the algorithm and the simulation analysis are given. Further practical application shows that the method satisfies the real-time and reliability requirements of TT&C data transmission.

## References

1. Zhai Z (1999) The data transmission rule in TT&C communication network. *J Spacecr TT&C Technol* 18(3):34–39(In Chinese)
2. Li W, Dang Q, Cheng G (2011) Adaptive method of double plane information non-loss transmission. *J Spacecr TT&C Technol* 30(2): 33–36(In Chinese)
3. Lin J, Zhang J (2011) Algorithm for network delay estimation based on end-to-end data moment. *Comput Eng* 37(10):32–35(In Chinese)
4. Ye X, Tse D (2006) Inference of link delay in communication network. *IEEE J Sel Areas Commun* 24(12):2235–2248
5. Fan J, Yang D, Shen Z, et al (2007) Simulation of network time delay based on UDP transport protocol. *Comput Simul* 24(8):102–106(In Chinese)
6. Gurewitz O, Cidon I, Sidi M (2006) One-way delay estimation using network-wide measurements. *IEEE Trans Inf Theory* 52(6):2710–2724
7. Hou C, Ding Y (2011) Improved 2-tuple linguistic model based data fusing method for evaluation. *J Syst Simul* 23(10):2142–2146(In Chinese)
8. Barzilay R, McKeown KR, Elhadad M (2005) Information fusion in the context of multi-document summarization. *Computat Linguist* 31(3):297–328

# Chapter 79

## Discovery of Static Test Configuration Model and Data Model Based on TTCN-3 Test Systems

Yongpo Liu, Shuangmei Liu, Ji Wu and Chuangye Chang

**Abstract** Aimed at the comprehensibility, reusability and maintainability, the thesis presents the reverse model recovery for the legacy code developed by TTCN-3. It can also help tester and maintainers to verify the test implement, etc. The thesis introduces the discovery of static test configuration model and data model based on the reverse model discovery system framework.

**Keywords** TTCN-3 · Reverse engineering · Test system · Static test configuration model · Data model

### 79.1 Introduction

With the development of TTCN-3, it had been used in many test fields. The test system based on TTCN-3 had many features as same as software development. But with the growth of the size of test systems and the alteration of testers, it had become increasingly difficult to manage and maintain the large-scale test system. So, by using reverse engineering based on TTCN-3, it can help testers design the test system from higher level and verify the consistence between designation and implementation. It is of great importance and value to maintain, extend and estimate the test systems [1].

In other paper named “The Designation of The System Framework of Reverse Model Discovery Based on TTCN-3 Test System”, the system framework of reverse model discovery had been designed and the static analyzer based on TTCN-3 had been also achieved by extending TRex. On this basis, the static test configuration model and data model were designed and implemented in this paper.

---

Y. Liu (✉) · S. Liu · J. Wu · C. Chang  
School of Computer Science and Engineering, Beihang University,  
Xueyuan Road 37, Haidian District, Beijing 100191, China  
e-mail: liuypo@sei.buaa.edu.cn

## 79.2 Discovery of Static Test Configuration Model

Test configuration is very important for a test system, it can give a description of the connectivity between test units and the number of test units. In the process of starting and executing a test, the testers can implement black-box testing by configuring a complete test scenario. In U2TP, the test configuration had been proposed, it is necessary to find test configuration information by reverse engineering, so that the information needed in a test could be unfolded in front of the users in a friendly way [2].

### 79.2.1 Extract Basic Information

It is very important to extract the basic information for reverse engineering. Defined in U2TP, the basic information of a test system is the test units and the relationships among them. Here only concerned to extract the basic information of test units. As TTCN-3 specification, the test units are declared as component. Every component has many ports to interact with the external world. The port is to be defined the way between message interacting, there are two ways, that is message-based or procedure-based. In a port many data formats are defined as record, record of, set, set of, union, enum and signature which are in TTCN-3. So, when traversing syntax trees, the keywords concerned in this paper are mainly ConstDef, SignatureDef, RecordOfDef, SetDef, SetOfDef, EnumDef, UnionDef, PortDef, ComponentDef.

According to the grammar, the information and relationships between symbols have been restructured in the syntax tree. There are two ways to operate the Node, as follows [3]:

- The first child can be gotten by `Node.getFirstchild()`, which includes detailed information.
- The second child can be gotten by `Node.getNextsibling()`, which mainly contains the possible adjacent elements in the syntactic description.

The whole information of a syntax element can be gotten by traversing the syntax tree. The syntax elements defined in TRex are shown from Figs. 79.1, 79.2, 79.3, 79.4, 79.5 and 79.6, which are often used in TTCN-3, and also concerned in this paper.

When extracting the basic information, many recursive algorithms will be used because of complex nested relations among these syntax elements.

The basic information is saved in TestArchitecture, which is a meta-model of a test system and defines test units and the relationships between them. Algorithm 1 is to extract the basic information. In order to describe conveniently, this part only introduces the information of test units. The configuration between units will be

Fig. 79.1 Record class

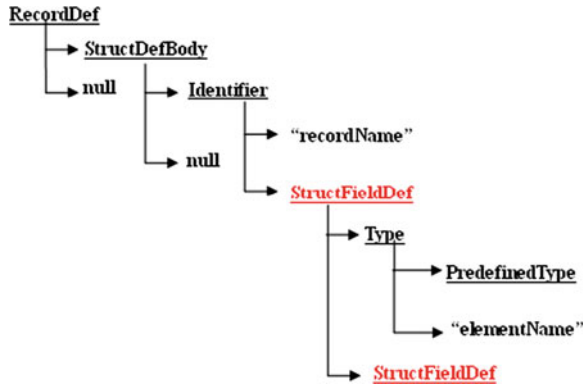


Fig. 79.2 Set class

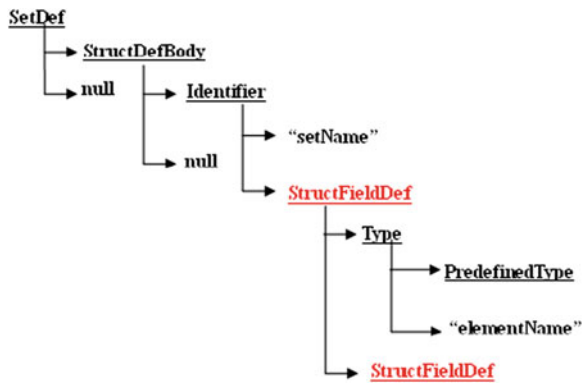
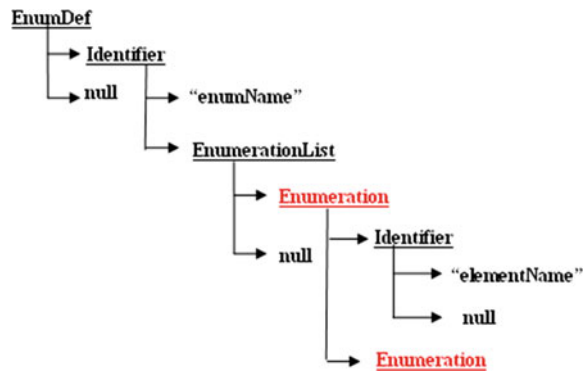


Fig. 79.3 Enum class



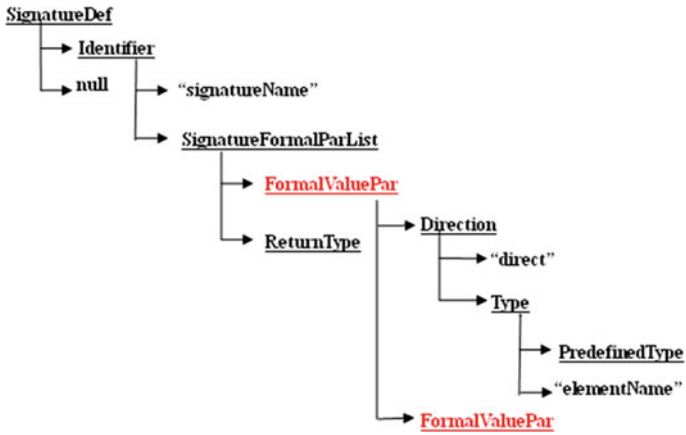


Fig. 79.4 Signature class

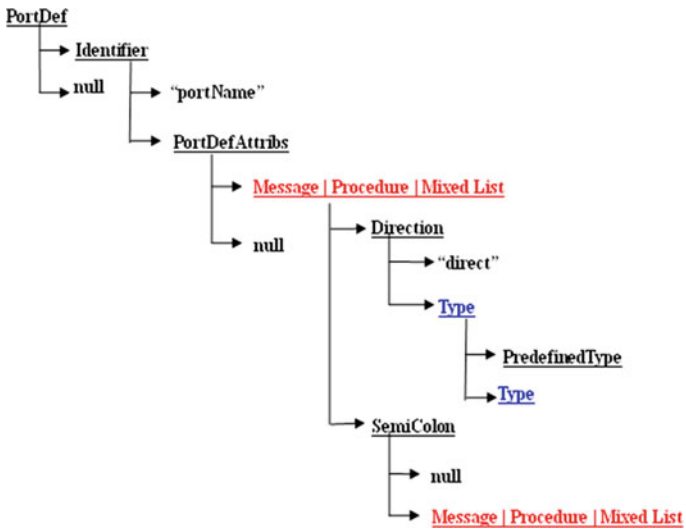


Fig. 79.5 Port class

discussed in detail in Sect. 2.2. In this paper, although Arbiter, SUT and Scheduler are designed, which are the necessary parts of a model, but not implemented because there is no one-to-one relation between TTCN-3 and the modules [4].



**Algorithm 1** Extracting the basic information

```

INPUT: Project := { filei | filei is the TTCN-3 script in the project, filei ∈ File }
      File := { module | module is the TTCN-3 module in the file }
OUTPUT: TestSystem := { <TestArchitecture, TestBehavior, TestData, Time> }
ALGORITHM BEGIN
TestSystem.name := Project.name;
FOR each file in the Project
  module := Module(file);
  FOR each type in module
    IF type == Structure ∨ Enumeration ∨ Signature
      TestArchitecture.DataType.add(type);
      IF type == Port
        TestArchitecture.Interface.add(type);
      IF type == Component
        TestArchitecture.TestComponent.add(type);
    END FOR
  FOR each template in module
    TestData.DataPool.add(template);
  END FOR
  FOR each function in module
    IF function == TestCase ∨ Function ∨ Altstep
      TestBehavior.Behavior.add(function);
    END FOR
  END FOR
TestSystem.add(TestArchitecture, TestBehavior, TestData, Time);
RETURN TestSystem;
ALGORITHM END

```

TestArchitecture meta-model is shown in Fig. 79.7. There is a fine mapping relation between *Algorithm 1* and the model.

### 79.2.2 Extract Static Configuration Information

By analyzing static abstract syntax tree and symbols, all the possible test configuration information can be found among all the test systems.

In a test system based on TTCN-3, the test configuration is often defined in the test function which has three kinds, such as testcase, function, and altstep. Here the test configuration and behavior statements are mainly concerned, which are in FunctionStatement. FunctionStatement mainly includes ConfigurationStatements,

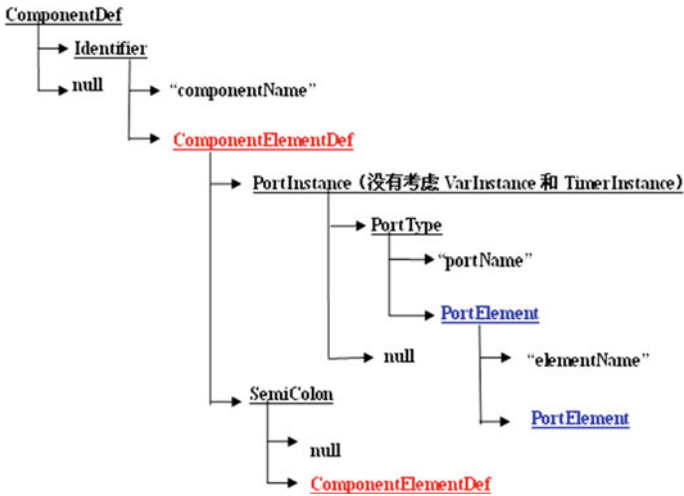


Fig. 79.6 Component class

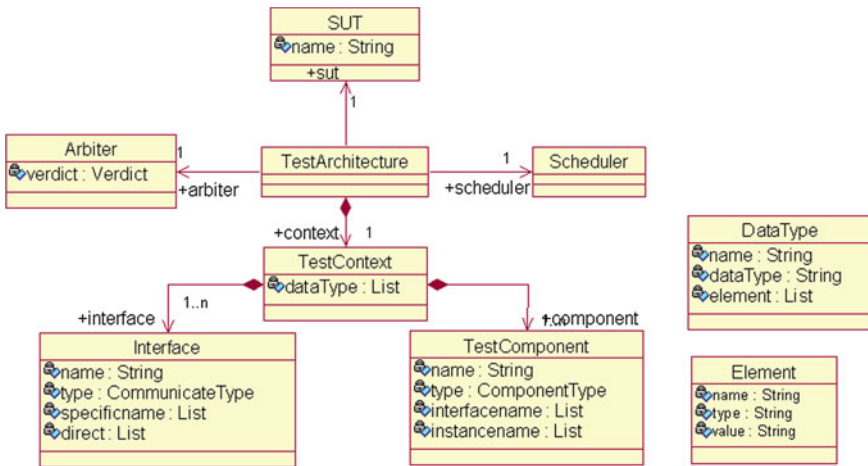


Fig. 79.7 The meta-model of TestArchitecture

CommunicationStatements, BasicStatements, BehaviorStatements and TimerStatements, shown as Fig. 79.8.

- ConfigurationStatements includes *ConnectionStatement*, *MapStatement*, *DisconnectStatement*, *UnMapStatement*, *DoneStatement* and so on.
- CommunicationStatements includes *SendStatement*, *CallStatement*, *ReplyStatement*, *RaiseStatement*, *ReceiveStatement*, *GetCallStatement* and so on.
- BasicStatements includes *Assignment*, *LogStatement*, *ConditionalConstruct*, *LoopConstruct*, *SelectCaseConstruct* and so on.

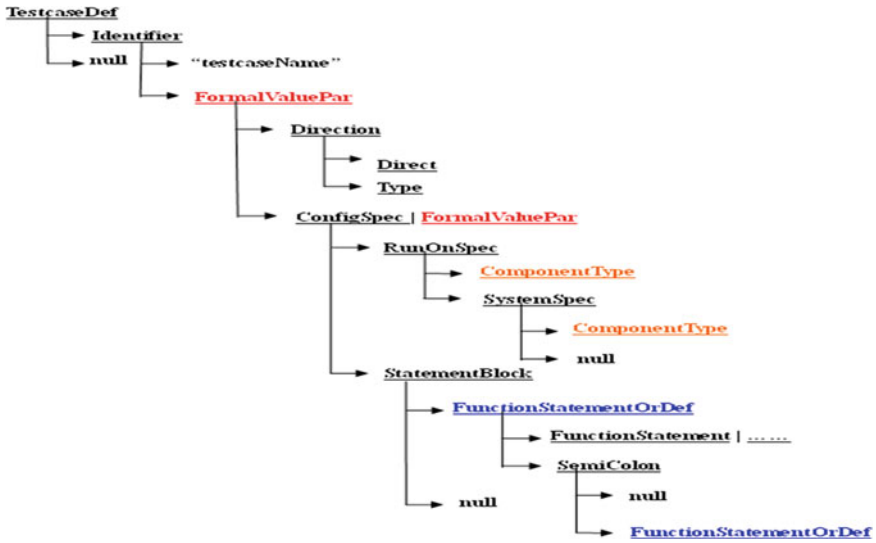


Fig. 79.8 Structure of Testcase

- BehaviorStatements includes *TestcaseInstance*, *FunctionInstance*, *AltstepInstance*, *ActivateOp*, *DeactivateStatement* and so on.
- TimerStatements includes *StartTimerStatement*, *StopTimerStatement*, *TimeoutStatement*.

The underlined declaration statements are mainly concerned in this paper. The call relations among test functions are shown as follows.

- There are two ways when testcase calls function, one is the direct call, when testcase and function are both running on the same test component; Another is to instantiate a test component that calls function running on the new instantiating test component, which is a PTC.
- There are also two ways when Function calls Function. One is the direct call, when they are running on the same test component; Another is to instantiate a test component which can call Function.
- Altstep can only be called by above functions by activate.

Algorithm 2 is to extract static test configuration information which are function call information and test configuration information of TTCN-3. The message interaction and the timer are not taken into in this section.

### Algorithm 2 Extracting the static configuration information

```

INPUT: Project := { filei | filei is the TTCN-3 script in the project, filei ∈ File }
      File := { module \ module is the TTCN-3 module in the file }
OUTPUT: TestBehavior := { behavior | behavior ∈ Behavior }
ALGORITHM BEGIN
FOR each file in the Project
  module := Module(file);
  FOR each function in module
    behavior.setName(function.Name);
behavior.setComponent(function.runsOn);
IF function == TestCase
  behavior.setType(TESTCASE);
IF function == Function
  behavior.setType(FUNCTION);
IF function == Altstep
  behavior.setType(ALTSTEP);
FOR each statement in function
  IF statement == MapState ∨ ConnectionState
behavior.setConfigInfor(statement);
  IF statement == FunctionInstance
behavior.setInnerCall(statement);
  END FOR
  TestBehavior.Behavior.add(behavior);
  END FOR
  END FOR
RETURN TestBehavior;
ALGORITHM END

```

Testers or maintainers can comprehend the test system quickly by using static configuration information, and can grasp the specific information of test configuration and scenario from every testcase. TestBehavior describes the function call and test configuration of TTCN-3, shown as Fig. 79.9.

It includes many behaviors to indicate the test functions of TTCN-3, such as testcase, function and alstep. Behavior records the information of name, type, component and so on from a function. Every behavior uses TestConfiguration to record the test configuration information used by itself and the relationship between function calls. The recording process is as follows [5]:

- In the definition statements of the test function, the parameters of Map or Connect between test units will be written into invoker, invokeprot, receiver, and receiveport, that is to record the test configuration information.

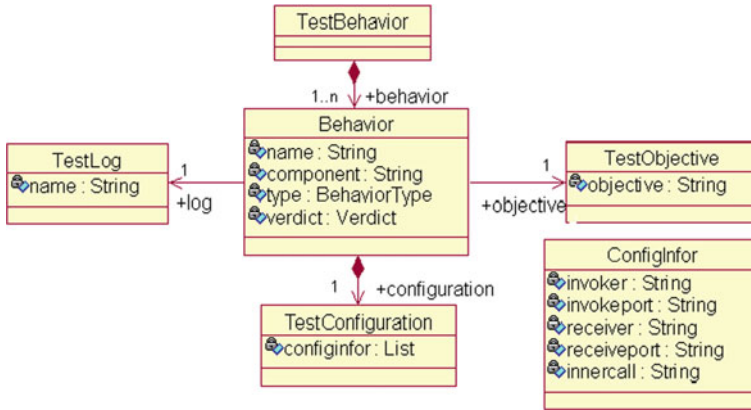


Fig. 79.9 The meta-model of TestBehavior

- In the definition statements, the name of called function will be written into innercall meta-model of ConfigInfor, in order to record function calling information.

### 79.3 Discovery of Data Model

In a test system, it is an important thing to save and partition test data. Effective organization and management for the test data not only can keep a very valuable test legacy, and higher levels of test requirements can also be achieved by using different data selection and combination strategy. Usually in a TTCN-3 test, testers need design and produce a large number of test data. For this reason, the U2TP specifically proposed the concept of test data, and data pool is one of the most important parts [6].

#### 79.3.1 Extract Data Model

TTCN-3 test specification provides rich types and mechanisms to define test data whose types are basic type, structure type and predefined type, and whose mechanisms are inheritance, reference and parametric. As a special test language, the test data and configuration of TTCN-3 are both realized by encoding. But in the course of testing, multiplexing and update the test data is very important, so it is necessary to have a good test data management system [6]. In this paper, by using reverse engineering, the test data defined in the script is structured by the way of

data pool in U2TP. In order to reversely find the data model oriented data pool, here the templates and data types are mainly analyzed, namely to analyze the TemplateDef structure in the abstract syntax tree, shown as Fig. 79.10.

- The template provides a way to organize and reuse test data, including simple inheritance relationship.
- The template can be parameterized, which makes it be seen as function, and be instantiated by parameter mechanism.
- The template allows matching mechanism, which is another strong function provided by TTCN-3, and eliminates the cumbersome comparison between return and expectations in the testing process.
- The template can be used in a message-based communication, also be used for process-based communication.

In this paper, the template inheritance, parametric and inline forms are not considered, and the template defined by users can be only found. Because the template allows to be nested in the defining process, this makes it difficult to show the template. So the method used in this paper is to remove the nested template from original structure, which is named a new template and then saved. Such that, the original structure retains only the new template reference, effectively solves the complexity of nested template. Finding data model helps testers or maintainers manage and maintain test data in a large scale.

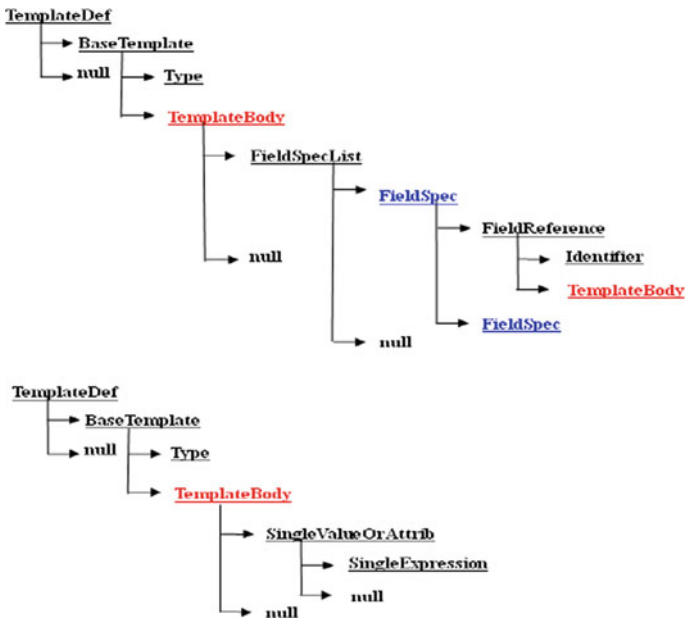


Fig. 79.10 Structure of template

### 79.3.2 Select and Partition Data Model

Although discovery of data model can show the test data of the test system to testers and maintainers in a graphical way, but it is loose and lacks of model representation [7]. By data model selection and partition, the test data of the test system can be shown with the form of model, which achieves better abstraction. The meta-model of test data is shown in Fig. 79.11.

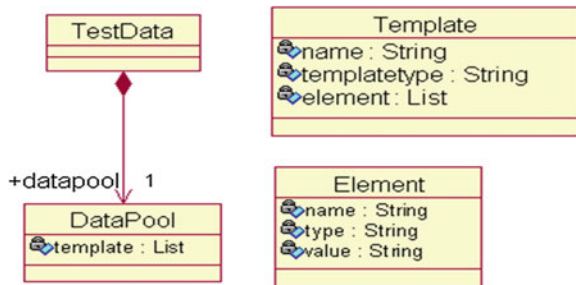
In the TTCN-3 test script, the data type is defined by type, while in the template the specific data instances of a data type are defined by type. So the DataPool meta-model of TestData is a set of templates. The Template meta-model records detailed information of a template, including name, type and value. In the DataPool meta-model, the selection and partition of test data can be achieved.

TTCN-3 has many types, each template is actually an instance of a data type, thus the data type can be selected following the template. In addition, the template can be partitioned by group keyword in TTCN-3 [7]. Each group actually represents a kind of test data with common properties (such as abnormal data). So you can partition the equivalence class to follow the template where the group is. In general, the template can be chosen according to the type, and can also be partitioned in accordance with equivalence class in TTCN-3 [8]. So after all the templates are identified, standards of selection and classification are given in this article as follows.

- For the templates that belong to the same type, in this article they will be classified as category so that testers can choose according to test data type, namely data selection;
- For the templates that belong to the same group, this article classifies them as a group, and provides a mechanism that allows testers to detect equivalence classes of test data distribution, namely data partition;

Based on the data model, TTCN-3 testers can achieve the higher coverage test in a way of permutation and combination by selecting the data from data pool; or achieve more high-level security test to vary the data from data pool; or can be separated from the specific test script to design test data, which avoids to modify the script code for maintaining test data.

Fig. 79.11 The meta-model of TestData



## 79.4 Conclusion

In this paper, static test configuration model and discovery of data model are introduced in detail on the basis of the framework of reverse model discovery system.

The test configuration and data plays an important role in a test system. Test configuration is used to describe the number of test components in a test system, and the connecting relationship between test components. Effectively organizing and managing the test data not only can save a valuable test legacy, but also can use different data selection and combination strategy to achieve higher levels of test requirements.

**Acknowledgments** This paper is contributed by National Key Laboratory of Software Development Environment Fund (SKLSDE-2010ZX-15) and High Technology Research and Development Program of China 863 Fund (2009AA01Z145).

## References

1. Jens G (2000) TTCN-3-A new test specification language for black-box testing of distributed systems. In: Proceedings of the 17th international conference and exposition on Testing Computer Software (TCS 2000), Theme: Testing technology vs. testers' requirements, (Washington, D.C.) pp 760–769
2. Peng S, Zhu Q (2006) Inverse modeling based on source code analysis. *Appl Res Comput* 7:52–54 (in Chinese)
3. Sun Ch, Zhou J (2005) A tool for automated software architecture recovery of javabeans-based applications. In: Proceedings of ASWEC 2005, IEEE computer society, Brisbane, Australia, pp 270–280, Mar 29–Apr 1
4. Dai ZR (2004) Model-driven testing with UML 2.0, second European workshop on model driven architecture (MDA) with an emphasis on methodologies and transformations (EWMDA'04), Canterbury, England, pp 648–655, Sept 2004
5. Din G (2004) A meta-model for TTCN-3. 1st international workshop on Integration of Testing Methodologies, ITM 2004, Toledo, Spain, pp 323–330, Oct 2004
6. TRex website [OL] (2006) <http://www.trex.informatik.uni-goettingen.de>
7. Paul B, Dominic E (2009) TRex—The refactoring and metrics tool for TTCN-3 test specifications, TAIC-PART. In: Proceedings of the testing: academic & industrial conference on practice and research techniques, 2010, pp 90–94
8. European Telecommunications Standards Institute (ETSI) (2007) The testing and test control notation version 3; Part 1: TTCN-3 core language. ETSI standard ES 201 873-1 V3.2.1, Mar 2007



# Chapter 80

## A New Method of Unknown Radar Signals Sorting

Xiaofeng Wang, Xuzhou Zhang, Runlan Tian and Xinglong Qi

**Abstract** Aimed at the problem of worse real-time in the sorting of unknown radar signals, a new signal sorting method based on the DBSCAN is put forward. Using the parameter distribution characteristics of radar pulse data, the method firstly searches a number of reference points to properly represent original data points, and then, make use of their density connected character to cluster the reference points. Because of the decrease of the data involved in clustering, the method overcomes the immense calculated quantities of original algorithm, and enhances the clustering speed of unknown radar signal sorting. Computer simulation results show that the proposed algorithm can effectively sort the unknown radar signals.

**Keywords** ECM · Signal sorting · Density-based clustering

### 80.1 Introduction

With the wide applications of radar in aircraft, ships, air defense systems and a variety of offensive weapons, and the increasing of electronic jamming equipment, the electronic surveillance has become very dense and complex [1]. In the known radar signal processing, the use of correlation between the priori knowledge and radar the received signal [2], sorting accuracy and real-time have reached a certain height. But the problem sorting unknown signal from the density radar pulse streams which contains a lot of interfering pulse has been restricting the radar surveillance data processing.

---

X. Wang (✉) · X. Zhang · R. Tian · X. Qi  
Department of Electronic Engineering, Naval Aeronautical Engineering Academy,  
No 188 ErMa Road, YaiTai City, China  
e-mail: wxf870516@126.com

Cluster analysis is an important technique in data mining. It can divided data into some categories based on the similarity between data objects [4]. We can use the special nature of cluster analysis to sort unknown radar signal. DBSCAN [5, 6] cluster is a density-based clustering algorithm, which does not require a priori information, adopt to the arbitrary shape clustering, and is not sensitive to the order of data input. The algorithm also has the ability to handle abnormal data, but is not suitable for heavy data processing because of its complexity. This paper first analyzes the DBSCAN clustering. Aiming at the complexity problem of DBSCAN, the paper use a certain number of reference points to represent a certain neighborhood range of data points, propose an improved algorithm which is applied to the unknown radar signal sorting. Simulation results verify the effectiveness of the improved algorithm.

## 80.2 DBSCAN Clustering

DBSCAN clustering has the advantage of no pre-determined the number of clusters, and DBSCAN is not sensitive to noise data, which solve the problem of noise pulses.

DBSCAN clustering determine the core object by querying the neighborhood  $\epsilon$  of each data in data set  $D$ , and then decide how to extend the class. So a large part of the algorithm is to execute the query region. The time complexity of DBSCAN is  $o(n \cdot n)$ , the time complexity of DBSCAN with the R \* tree is  $o(n \cdot \log n)$ , where  $n$  is the total amount of data. With the data points increase, DBSCAN clustering computation complexity increase as exception, so it is not suitable for handling large amounts of data.

## 80.3 Improved DBSCAN Clustering

As the computation complexity of DBSACN, it is difficult to meet real-time in radar signal sorting, this paper proposes an improved DBSCAN algorithm. The basic idea of improved algorithm is: a certain number of reference points are used to represent a certain neighborhood range of data points, establishment corresponding relationship between data point and the reference point, removal noise pulse based on the density of the reference point, and finally complete the reference point of clustering using breadth-first search algorithm, and the reference point and the corresponding relationship between the data points on the final completion clustering of data points.

### ***80.3.1 Basic Concepts***

Definition 1 point density: select the any point  $p$  at as the center in the data set, the number of data in the region radius as the point density, denoted by  $p_d$ .

Definition 2 reference points: select any point  $p$  and the radius  $R$  in the data set  $D$ , if met  $p_d > Th$ , called point  $p$  as a reference point, denoted by  $r$ , which  $Th$  is the density threshold. Note that the reference point is not a data point in data set, but a virtual point.

Definition 3 representative region: the region which reference point is  $r$  and radius is  $R$  is called the representative region, reference point can represent the data points in this region.

Definition 4 connected region: If the representative region of two reference points  $r_1, r_2$  is intersection, tangent, or overlap, that is the distance between  $r_1$  and  $r_2$  is less than or equal to 2 times  $R$ , we called representative region of two reference points  $r_1, r_2$  is connectivity.

### ***80.3.2 Find the Reference Point***

Select the first data  $p_1$  in data set as a reference point  $r_1$ , calculate the distance  $d$  between data points  $p_2$  and reference point  $r_1$ . If  $d \leq R$ , the information  $p_2$  will be added to the reference point, update reference point  $r_1$  base on mean of all dimensions parameter values. If  $d > R$ ,  $p_2$  does not belong to the representative region of the reference point  $r_1$ , establish a new reference point use dimensional parameters  $p_2$ . Count the distance between next data points and all the reference point until complete all of the data points.

### ***80.3.3 Establish the Corresponding Relationship Between the Reference Points and the Data Points***

There is only one data point  $p$  which the distance is less than or equal to  $R$ , establish the correspondence between the reference point and data point  $p$ . when the distance are less than or equal to  $R$ , we can prove that the represent region of these reference points are connectivity, that is to say they belongs to same class, this paper will establish correspondence with the nearest reference point. When all data points are established with the reference point, the number of representative region is the density of the reference point.

Compare with the density of all the reference points with density threshold  $Th$ , filter out the noise density of less than the threshold reference point, the remaining reference point is the non-noise data reference point.

### 80.3.4 Cluster the Reference Points

We can see from the definition 4, when the representative region of two reference points  $r_1$ ,  $r_2$  is intersection, tangent, or overlap, the representative region is connected. Clustering the reference point is to find these connected regions and put these connected regions into a class. So we can use the Breadth-first search algorithm to complete the clustering of reference points. Finally, according to the corresponding relationship between the data points and the reference points, and the clustering results of reference points, we can establish correspondence between the clustering and data points.

Among them, the selection of parameter  $R$  has a greater effect to the clustering result, the little with  $R$ , the lower the efficiency of the algorithm, when the  $R$  is too little, all of the data point will be regarded as the reference points to cluster. When the  $R$  is too big, the algorithm cluster all data points to one cluster. According to the experimental results, when the parameter  $R$  is taken from 0.005 to 0.009, the unknown radar signal sorting can be accomplished perfectly.

In the radar signal sorting, the number of noise signal is much smaller than the real radar signal, so the density of the noise signal is less than the density of reference point. The function of density threshold effect is to filter the noise signal, so we take data and un-filter noise ratio as the density threshold. That is to say the average density before filtering noise can effectively filter out the noise reference point.

### 80.3.5 Algorithm Steps

The steps of radar signal clustering sorting method are shown in Fig. 80.1.

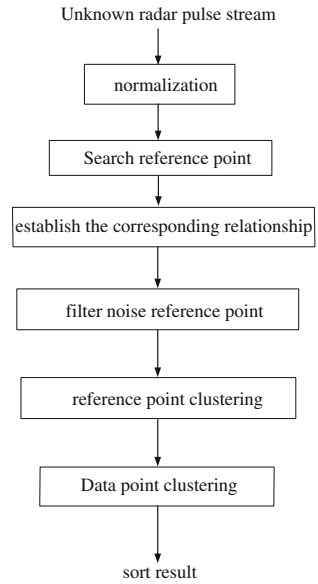
## 80.4 Simulation Analysis

Select the carrier frequency (RF), pulse width (PW) and the direction of arrival (DOA) as the characteristic parameters, three attributes of information constituted a characterization vector [8]. Add 30 % of the total radar pulses number noise pulse. The specific pulse parameters are shown in Table 80.1. All simulations are with MATLAB.

Using the improved DBSCAN algorithm cluster pulse data in Table 80.1, parameters  $R$  obtained 0.008. According to the following formula, we can obtain the clustering results shown in Table 80.2.

$$\text{sorting accuracy} = \frac{\text{accuracy sorted pulse number}}{\text{existent pulse number}} \times 100\%$$

**Fig. 80.1** Algorithm steps

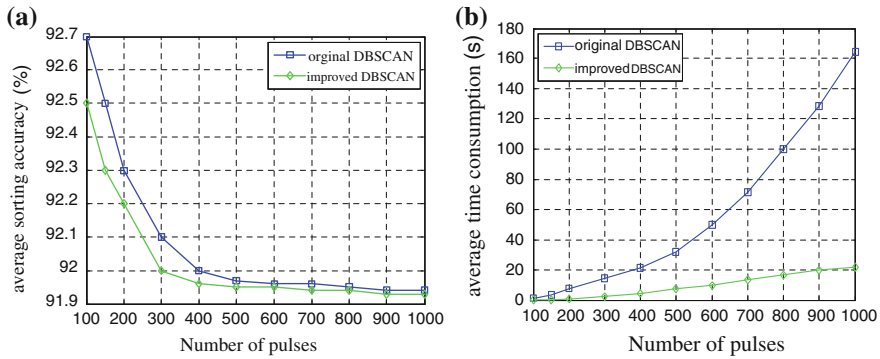


**Table 80.1** The radar pulse simulation parameters

Signal	RF/GHz	PW/ $\mu$ s	DOA/( $^{\circ}$ )	PRI/ $\mu$ s	Number
Radar 1	8.9–9.3 RF agile	0.3–0.4	38–41	30–40 PRI staggering	120
Radar 2	9.0–9.3 RF hopping	0.8–0.9	36–39	30–60 PRI jittering	100
Radar 3	9.4–9.8 RF agile	0.4–0.5	40–43	40–45 PRI jittering	115
Radar 4	9.6–9.8 RF hopping	0.9–1.0	42–45	40–70 PRI staggering	90
Noise	8.9–9.8	0.3–1.0	36–45	–	127

**Table 80.2** Sorting the results of improved DBSCAN clustering method

Radar serial number	1	2	3	4
Existent pulse number	120	100	115	90
Sorted pulse number	125	101	113	88
Accuracy sorted pulse number	106	90	110	84
Error sorted pulse number	9	7	0	0
Omitting sorted pulse number	14	10	5	6
Noise number of sort result	3	4	3	2
Sorting accuracy (%)	88.3	90.0	95.6	93.3



**Fig. 80.2** Comparison chart of algorithm performance. **a** average sorting accuracy, **b** average time consumption

Simulation results show that four of the radar signal can be sorted correctly, and effectively eliminate the impact of noise pulses. In order to verify the performance of the algorithm, we used improve DBSCAN clustering and the original DBSCAN clustering to sort the same set of radar pulses parameters of. The average accuracy and average consumption of the two sorting algorithms are shown in Fig. 80.2. The simulation results show that: with the increase in the amount of data, the average accuracy rate of improved sorting algorithm declined slightly compared to the original algorithm, the time consumption of original DBSCAN clustering of increases exponentially, but the improved DBSCAN clustering has approximate linear time complexity, which is applied to sort high density signal.

To further validate the effectiveness of the algorithm in this paper using in radar signal sorting, we respectively use k-means clustering and SVC clustering to process data in Table 80.1 Three experiment all have 100 Monte-Carlo test, which is  $k = 4$  in the k-means clustering, the initial centers randomly selected to run in, SVC is under optimal parameters condition. Simulation results comparison of three methods are shown in Table 80.3.

Simulation results show that: the average time consumption of this paper method slightly more than k-means clustering, but the average sorting accuracy is 16.3 % more than k-means clustering. Compared with the SVC clustering, the average sorting accuracy of this paper method was slightly lower. However, SVC clustering id not suitable for handling a large number of radar data because of its severe time consumption. Although the time complexity of k-means clustering is

**Table 80.3** Comparison of three methods

Method	Average sorting accuracy (%)	Average time consumption (s)
k-means	75.4	4.25
SVC clustering	92.5	215.7
Improved DBSCAN clustering	91.7	7.62

low, but k-means clustering requires pre-specified k value and the initial center of clustering, its results depend heavily on prior knowledge, and it is sensitivity to noise pulses, therefore it not suitable for unknown radar signal sorting. Therefore, this paper method has higher overall performance in sorting unknown radar signals.

## 80.5 Conclusion

This paper proposes a radar signal sorting method based on improved DBSCAN clustering, respectively compare with the original DBSCAN clustering, k-means clustering and SVC clustering, validate the proposed algorithm can complete the unknown radar signals sorting in a shorter time, which has a good generalization performance.

## References

1. Li H, Yu H, Cai Y, Tao R (2005) Overview of the crucial technology research for radar signal sorting. *Syst Eng Electr* 27(12):2036–2039 (in Chinese)
2. Hassan HE (2003) Joint deinterleaving/recognition of radar pulses. In: *Proceedings of the international radar conference 2003*, pp 171–181
3. Guo J, Chen J (2006) Clustering approach for deinterleaving unknown radar signa. *Syst Eng Electr* 28(6):854–863 (in Chinese)
4. Jia WH, Micheline K (2002) *Data mining*. Higher Education Press, Beijing (in Chinese)
5. Qian W, Gong X, Zhou A (2003) Clustering in very large databases based on distance and density. *J Comp Sci Technol* 18(1):67–76 (in Chinese)
6. Bian F, Huang M (2009) A grid and density based fast spatial clustering algorithm. *Comput Soc* 23(3):132–135 (in Chinese)
7. Januzaj E, Kriecel H-P, Pfeifle M (2004) Density based distributed clustering. *Proceedings of the 9th international conference on extending database technology*
8. Guo Q, Wang C, Guo L (2008) Application of segment clustering in radar signal sorting. *J Beijing Univ Posts Telecommun* 31(2):132–135 (in Chinese)

# Chapter 81

## The Creation of Interactive Three-Dimensional Microscope Based on VRML and JavaScript

Di Wu

**Abstract** When constructing sophisticated 3D interactive apparatus with VRML, a method is usually adopted that is manipulating another object with a selected one. Therefore, the way of achieving the linkage movement of three-dimensional objects on website was crucial. Adjustment of the apparatus for complex interaction can not simply be operated with VRML itself. This article takes the creation of interactive three-dimensional microscope for example, with the technique of JavaScript programming which realized the interactive operation adjusting the place of microscope lens horizontally and vertically, a complex interaction and coordinated controlling campaign achieved.

**Keywords** VRML · JavaScript · Interactive · Microscope

### 81.1 Introduction

At present, students can log on to their college physical experiment Web site to study the principle and the operation of experiments using their spare time. But the problem is that the current web site with mostly describing text and the displaying pictures of apparatus or animation of the experiment with the process being set in advance by the programmers. This is far to the actual experiment. Virtual Reality (VR) is a three-dimensional environment simulated by computer, the user can enter the environment through the computer and can manipulate the system and interact with the objects [1–3]. VRML is the second generation of the key technologies on the Web [4–8]. It makes the possibility to achieve three-dimensional virtual scene. Web browsing with transmitting very little data in the Internet

---

D. Wu (✉)

College of Physical Science and Technology, Dalian University, No. 10 Xuefu Street,  
Dalian Economic and Tech. Development Zone, Dalian 116622, China  
e-mail: wudi@dlu.edu.cn



overcomes the shortcomings of the general three-dimensional simulation software. Therefore using VRML to develop 3D virtual experiments to solve these problems should be a better way [9–12].

But VRML itself does not perform complex interactions, and thus for the complex interactions such as the adjustment of apparatus can not be carried out only by using VRML itself. The linkage complex motion controlling method of a virtual experiment was studied using JavaScript in this paper [13, 14]; Eventually, an interactive 3D experimental environment capable for usage of microscope was established.

## 81.2 Linkage Control Principle

An interaction in virtual environment can be achieved through the node. Changes in the state of VRML will be recorded as an “event”, the node induced the changes of other nodes issue “eventOut”, the changed node accept the “eventIn”. VRML provides a variety of sensor nodes which monitor some user actions like clicking on some object, getting close to some object etc. It includes TouchSensor node and CylinderSensor node etc. The TouchSensor node can change the procedures such as animations, sounds, changing colors, etc. when mouse clicks on object and then makes a corresponding physical actions; The CylinderSensor node can move the mouse into a shape around the axis of rotation; Data exchanging among nodes in VRML can control objects in the scene forming some simple linkage action: If turn on the switch a lamp lit and if approach the automatic doors the door opened. A node can send and receive events. ROUTES are a simple way of defining a path between an event generated by a node and a node receiving an event. Through the route and sensor nodes are basically able to achieve a simple linkage in the scene. However, it can not execute the complex movements such as the animated feature etc. VRML uses Script node to achieve this complex linkage controlling motion. Script nodes in the domain use JavaScript to achieve a specific period of the event handling, and then return the results of the calculation submitted to the Script nodes in the eventOut, and then through the routes pass the event to the scene of a relevant field to achieve the purpose of changing the scene.

## 81.3 The Key Program Design

3DMAX or MAYA can be used to build a three dimensional microscope model and then convert to VRML format file or use VRML directly to the construction of the three-dimensional model such as the experiment interface shown in Fig. 81.1. VRML allows the heavier of user to conduct real-time action on the scene, the scene node receives event through the eventIn, sends event through the issuing eventOut. The event system formed by one eventOut node to another eventIn node

**Fig. 81.1** Seat-type light source



makes VRML have the ability of dynamic interaction. To achieve the complex linkage functions the interactive function completed by VRML language itself is limited; it also needs the help of script. Script node allows users to create their own domain (field) and events, it supports multiple scripting languages such as VRMLScript, or JavaScript etc.

### ***81.3.1 Controlling of the Sodium Light Turn On and Off***

If press the switch ‘Switch Button’, sodium lamp lights yellow as shown in Fig. 81.1, means the sodium light turns on; if press again, the light turns off. This function is realized through the touch sensor ‘TouchSensor’. First define the sodium light, then define the switch, at last create a script to Change the lighting material parameters ‘emissiveColor’. The route map for the controlling process is shown in Fig. 81.2. When click on the touch sensor ‘ts’, the ‘isActive’ will be activated, the value passed to IFTOUCH function in the script ‘switch’ by ROUTE statement, the lflag value is set to 110 represents yellow, this value transmit to ‘imissiveColor’ parameters of sodium lights mm by ROUTE statement to turn on the sodium light; when click the touch sensor ts again, lflag set the original value and the sodium light turn off.

**Fig. 81.2** Routing map of the sodium light controlling system



**81.3.1.1 Definition of Lights**

The Program is as follows:

```

Transform {
  translation -0.6 1.2 0
  children [
    Shape {
      appearance Appearance {
        materialDEF mm Material {
          diffuseColor .6 .6 .6
          emissiveColor .1 .1 .1
          transparency 0 } }
      geometry Box { size 0.1 0.5 0.2 }
    }
  ]
}
    
```

**81.3.1.2 Definition of Switch**

```

Transform {
  translation 1.1 0.2 3.65
  rotation 1 0 0 1.57
  children [
    Shape {
      appearance Appearance {
        material Material {
          #defination of materials } }
      geometry Cylinder { radius 0.5 height 0.2 }
    }
    DEF ts TouchSensor { }
  ]
}
    
```

### 81.3.1.3 Controlling Script of the Light Switch

```

DEF switch Script {
eventIn SFBool IFTOUCH
eventOut SFColor lflag
field SFBool Check FALSE
url "javascript: function IFTOUCH(butt, eventTime)
{
    lflag=new MFColor();
    if(butt==true)
        {if(Check==FALSE)
            {lflag[0]=1.0;lflag[1]=1.0;lflag[2] =0.0;Check=TRUE;}
        else
            {lflag[0] =0.0;lflag[1] =0.0;lflag[2] =0.0;Check=FALSE;}
        }}"
ROUTE ts.isActive TO switch.IFTOUCH
ROUTE switch.lflag TO mm.emissiveColor .

```

## 81.3.2 Controlling of Microscope Lens Horizontal Movement

The horizontal movement of microscope lens implement through cylindrical sensor 'CylinderSensor'. First define the CylinderSensor the sensor bangtdgulun within the micrometer drum, then creating variable 'pingyis' script to realize the function. When click the cylindrical sensor bangtdgulun, the value of parameter 'rotation\_changed' will be transmitted to micrometer drumthrough by ROUTE making it rotation, Meanwhile, the changed value is also given to variable myrotation in script pingyis, through statement weiyi [0] = rotation.angle/3.14, change the angle of rotation into a translation distance and then attached it to the variable weiyi, this displacement variable is assigned to the variables 'translation' of the microscope lens to operate horizontally by routing statement ROUTE. The program listed below.

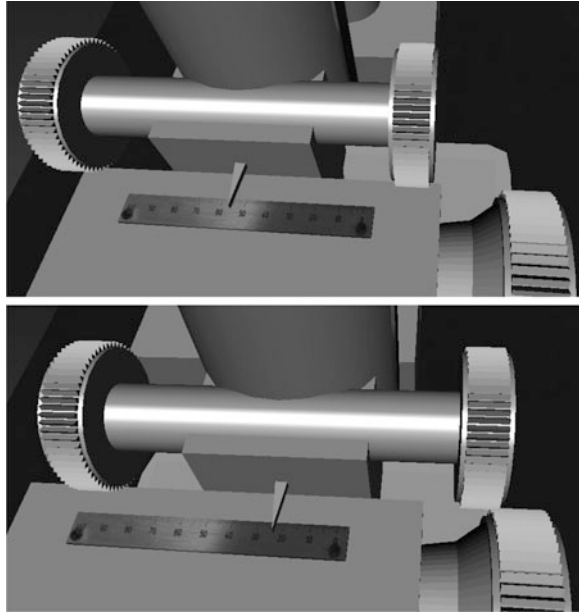
### 81.3.2.1 Definition of CylinderSensor

```

Transform {
    rotation 0 0 1 1.57
    children [
        DEF gulun Transform {
            translation 2 0.8 0.3
        children [
            Inline {url "jilun1.wrl" } }}
            DEF bangtdgulun CylinderSensor {
                maxAngle 1.57
                minAngle -1.57
            }}
    ]}

```

**Fig. 81.3** Movement result pointed on the ruler to the microscope head along horizontal direction



### 81.3.2.2 Controlling Script

```

DEF s Script {
  eventIn SFRotation    myrotation
  eventOut SFVec3f      weiyi
  url "javascript:
    function myrotation(rotation){
      weiyi[1]= rotation.angle/3.14;
    }";
  ROUTE bangtdgulun.rotation_changed TO gulun.rotation
  ROUTE gulun.rotation TO pingyis.myrotation
  ROUTE pingyis.weiyi TO pingyi.translation

```

The running profile is shown in Fig. 81.3.

### 81.3.3 Controlling of Microscope Lens Vertical Movement

The vertical movement controlling program is similar to the horizontal one. The scripts are as follows:

```

DEF pingyis Script {
    eventIn SFRotation myrotation
    eventOut SFVec3f weiyi
    url "javascript:
        function myrotation(rotation){
            weiyi[0]= rotation.angle/3.14;
        },}
    ROUTE bangtd.rotation_changed TO bang1.rotation
    ROUTE bangtd.rotation_changed TO bang2.rotation
    ROUTE bang1.rotation TO s.myrotation
    ROUTE s.weiyi TO cy.translation

```

### ***81.3.4 Build Up the Geometry Structure of the Experimental Apparatus***

The sodium light and model could be built up by VRMLPad or vrbuild [15]. Sodium light has three translucent light window, first we define one of them, then use 'USE' statement create others; Create a sodium light power model, switch has been designed in order to control the lights turn on and off; Create the three-dimensional structure of the microscope model, including eyepiece, objective lens, vertical and horizontal movements adjusting wheel for microscope lens, a ruler and other components etc.

## **81.4 Summary**

The coordinated controlling for the complex motion need the help of powerful high-level language to implement in order to compensate the deficiencies, such as logic determination and scene precise controlling of VRML its own. In this paper, the virtual experiment linkage animation technology was focused on by programming the Script node in VRML-based research, interactive 3D microscope and complete linkage has been designed. It will play an important role in the future teaching reform for injecting new vitality to the traditional teaching. VRML Web virtual technologies can be expected throughout the experimental teaching, designing the virtual products etc. It will be having a wide range of application prospects.

**Acknowledgments** The authors gratefully acknowledge the financial support of the Ordinary Higher Education Undergraduate Teaching Reform Research Project of Liaoning Province in 2012 (No.937).

## References

1. Zhang TJ, Ma ZJ, Jiang J, Pan RK (2007) Application of 3D animations in solid state physics teaching. *J Shenyang Norm Univ (Nat Sci Ed)* 25(4):454–456 (in Chinese)
2. Meng YT, Fu GS, Chang X (2009) Study of the interactive mechanism and its realization in VRML. *Electron Sci Tech* 22(6):40–44 (in Chinese)
3. Cao YH, Fan RQ, Fan JB, Ma TH, Zu J (2010) Study on reappearance technology of position and time for projectile based on virtual reality. *J Proj Rockets Missiles Guid* 30(1):91–94 (in Chinese)
4. Huang WL, Lu BH, Yang ZG, Jing N (2003) Introduction and application of VRML language. China Railway Publishing House, Beijing (in Chinese)
5. Wu BX (2004) Virtual reality modeling language VRML. Higher Education Press, Beijing (in Chinese)
6. Lu CH (2003) Introduction and enhancement of VRML. Peking University Press, Beijing, pp 201–202 (in Chinese)
7. Gavin B, Anthony P, Mark P (2012) Historical—VRML 1.0 specification. <http://www.web3d.org/x3d/specifications/vrml/>
8. McKenzie D (2012) Getting started with VRML. <http://www.vrmlsite.com>
9. Gu DH, Wang JJ, Yang L (2010) The research of electrical and electronic experiment teaching reform based on virtual reality technology. *China Mod Educ Equip* 13:55–3 (in Chinese)
10. Yu DX, Du JX (2010) Design and development of long-distance virtual teaching system based on VRML and JAVA. *Henan Sci* 28(5):600–604 (in Chinese)
11. Duan XY, Jiang P (2011) Exploration to virtual reality 3D key techniques based on cosmo browser plug-in. *Int J Digit Content Technol Appl* 5(6):391–402 (in Chinese)
12. Wu D, Wu YW, Liu J (2012) Construction of interactive 3D virtual air-cushion track experiment based on VRML and JavaScript. *Phys Exp* 32(6):17–20 (in Chinese)
13. Wang HP (2006) VRML elevation grid nodes based on the complexity of modeling. *Micro Comput Inf* 7(22):211–212 (in Chinese)
14. Wang JZ, Li DG, Zhang JC (2006) Engineering thermophysical experiment teaching system based on VRML. *Lab Res Explor* 25(12):1530–1532 (in Chinese)
15. Ligos Corporation (2010) V-Realm™ builder user's guide and reference. <http://www.few.vu.nl/~eliens/documents/vrml/V-Realm/>. Accessed 2 Oct 2010

# Chapter 82

## Research and Implementation of Face Detection System on Android Smart Phone

Xin Li, Yumei Zhai and Xiong Li

**Abstract** In order to improve the effect of detecting faces in Android smart phone, we proposed an effective face detection algorithm. Cluster analysis is used to segment skin region from the input color image in YCbCr color space. After the normalization of candidate regions being finished, the weighted Euclid distance is calculated between the reference templates and candidate regions. Based on this algorithm we implemented a face detection system on Android platform. Experiment proves that the system allows robust and has a high detection rate.

**Keywords** Face detection · Android · Cluster analysis · Template matching

### 82.1 Introduction

Face detection is one kind of technology that accurately confirms and locates faces which may exist in the input images. Android smart phone is developed on Android platform, an open source and popular operating system released by Google. Due to the combination with a lot of Google application software, the industry of Android smart phone develops rapidly. At the same time, people are paying more and more attention to the security of mobile phones. So the application of face detection and recognition embed in smart phone becomes more and more important. In recent years, face detection technology has become a research focus in the field of digital image processing and pattern recognition. At present

---

X. Li (✉) · Y. Zhai · X. Li  
College of Information Science and Engineering, Guilin University of Technology, Guilin,  
Guangxi, China  
e-mail: lx1996@glut.edu.cn

X. Li  
Guangxi Key Laboratory of Spatial Information and Geomatics, Guilin University of  
Technology, No.12 Jian Gan Road, Guilin, Guangxi, China



there are several main detection methods, such as neural network-based method [1], template matching method [2], support vector machines method [3], eigen-face-based method [4], knowledge-based method [5], and appearance-based method [6]. However, the methods above are not effective enough for complex face detection. In this paper, we proposed an algorithm combining skin-color segmentation with template matching. The face detection is implemented on Android smart phone giving fairly good results. Experimental results show that detection effect of this system is satisfactory and rate of detection comes up to 90.74 %.

This paper was organized as follows: Firstly, face detection algorithm is introduced and explained in Sect. 82.2. Secondly the experimental results are given with data and pictures in Sect. 82.3. Finally we draw our conclusion and put forward further enhancements in Sect. 82.4.

## 82.2 The Design of Face Detection Algorithm

### 82.2.1 Cluster Analysis-Based Skin Segmentation

At present, there are several popular color spaces [7], such as RGB, YUV, HSI, YIQ, YCbCr, etc. In this paper we adopt HSI color space, in which cluster feature of skin is perfect. The input face image is the formula transformed from RGB to HSI color space according to the following formula:

$$\begin{cases} Y = 0.299R + 0.587G + 0.114B \\ Cb = B - Y \\ Cr = R - Y \end{cases} \quad (82.1)$$

This paper takes sample pixels from 60 images downloaded from Internet. Skin pixels possess different cluster feature and frequency distribution in H-S subspace based on different ranges of Y. So we divide Y into eight ranges and respectively account frequency distribution of sample pixels in Cb-Cr subspace. Then, whether sample pixels belong to a region of candidate region or not depends on the set threshold. Cluster analysis-based skin segmentation method in YCbCr color space is expressed as follows:

- (1) Initialize: I and O respectively represent input and output pixelarray. Let  $m_i$  be the total of pixels in the range  $i$  of Y. Let T1 and T2 be the minimum and maximum threshold value.
- (2) for ( $i = 1$ ;  $i \leq 16$ ;  $i++$ ).
- (3) for ( $j = 1$ ;  $j \leq m_i$ ;  $j++$ ).
- (4) do  $F \leftarrow \text{FREQUENCY}(I[j])$ .
- (5) if  $T1 \leq F \leq T2$ .
- (6) then  $O[j] \leftarrow 1$ .

(7) else  $O[j] \leftarrow 0$ .

(8) return  $O$ .

After that, we smooth the skin detection results in order to eliminate most of the noise.

### 82.2.2 Normalization Processing

As candidate face regions are vulnerable to the variation of gray scale in template matching, we make normalization before calculating the similarity. Firstly, we cut out face region in the candidate face image according to the size of the reference template. Secondly, we calculate rotation angle of the candidate face using (82.2) and adjust the two-value image to get frontal faces.

$$q = \frac{1}{2} \arctan \frac{q}{p - r} \quad (82.2)$$

where

$$p = \sum_{i=1}^m \sum_{j=1}^n (x - \bar{x})^2 A[i, j] \quad (82.3)$$

$$q = \sum_{i=1}^m \sum_{j=1}^n (x - \bar{x})(y - \bar{y}) A[i, j] \quad (82.4)$$

$$r = \sum_{i=1}^m \sum_{j=1}^n (y - \bar{y})^2 A[i, j] \quad (82.5)$$

$$\bar{x} = \frac{1}{S} \sum_{i=1}^m \sum_{j=1}^n i A[i, j] \quad (82.6)$$

$$\bar{y} = \frac{1}{S} \sum_{i=1}^m \sum_{j=1}^n j A[i, j] \quad (82.7)$$

That  $(x, y)$  represents the coordinate of pixel  $x$  ( $i, j$ ).  $S$  indicates the area of candidate region.  $A$  is  $m \times n$  dimensional matrix of the candidate face region. Finally, we use formula (82.3) to normalize gray scale.

$$g(x, y) = b^{(cf(x,y)-a)} \quad (82.8)$$

where  $a, b, c$  are parameters required in histogram equalization process. And  $f(x, y)$  and  $g(x, y)$  respectively represent the image before and after normalization.

### 82.2.3 Template Matching

In this paper, we proposed a new method- weighted Euclid distance [8] to calculate similarity between reference template and candidate faces. In consideration of timeliness and efficiency required by Android smartphone, the size of the reference template should not be too large or too small. After testing, we choose reference template of  $65 \times 85$  size. In this paper the weighted Euclid distance is expressed as follows:

$$D(A, B) = \sqrt{\sum_{i=1}^m \sum_{j=1}^n \alpha_{ij} (A_{ij} - B_{ij})^2} \quad (82.9)$$

where A and B are the gray matrix of the candidate face and the reference template. Also

$$\alpha_{pq} = \frac{m \times n}{\sum_{i=1}^m \sum_{j=1}^n (A_{ij} - \bar{B})^2} \quad (82.10)$$

$$\bar{B} = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n B_{ij} \quad (82.11)$$

The small the weighted Euclid distance is the most candidate face regions are similar to the reference template. If the distance is smaller than the given threshold, the corresponding candidate face window is added to the face region.

Suppose that there are N windows in the candidate face region, the number of current detected window is w. Dw (A, B) represents current distance. Template matching method based on weighted Euclid distance is expressed as follows:

- (1) Initialize: Let T be the threshold of distance.
- (2) for (w = 1; w ≤ N; w ++).
- (3) calculate the mean gray of the reference template using formula (82.11); calculate Dw (A, B) using formula (82.9).
- (4) if DW (A, B) < T then the window is retained as a face region.
- (5) else the window is retained as non-face region.
- (6) w ← w + 1.

## 82.3 Experimental Results

### 82.3.1 Development Environment and System Implementation

This system is an intelligent project which was developed using Java in eclipse3.5 environment and tested on Android smartphone. We installed JDK 1.6 and

Android SDK 2.3, and configured the environment variables on Windows XP operating system with Pentium Dual-Core T2330 processor and 1G memory. Then, Eclipse3.5 was installed and started. After that, we created an Android emulator.

We implement this system in eclipse3.5 integrated development environment by executing Java application and calling Matlab-based face detection function. The face detection algorithm was coded by Matlab R2009a, and each detection function is encapsulated as a method of Java class. Matlab and Java communicate with each other through JNI (Java Native Interface). Android application is executed by calling Matlab methods.

### 82.3.2 Experimental Results and Analysis

In order to test the performance of the system, we choose 180 color images collected from Internet, which are classified as single face test set and multiple faces test set. The two test sets are expressed as follows:

(1) Single face set consists of 80 single-face color images, whose collection methods are shown in Table 82.1.

(2) Multiple faces set consists of 100 color images, which contain 326 faces. The collection parameters of them are the same as single face set.

76 faces are detected and 4 missed in the single-face test set, the correct detection rate of which comes up to 93.75 %. By contrast, we can see that experiments have achieved a detection rate of 87.73 % in multiple faces set, among which 284 faces are detected and 32 faces missed. Some experimental results of two test sets are shown in Figs. 82.1 and 82.2.

Although some of images are partially occluded and rotated, and there exist lands and walls which are similar to skin color, the system can still successfully detect most of the faces. As shown in Fig. 82.2, one face is missed in the fourth image, the reason for which is the face is shaded with a pair of sunglasses and a hat. The last image. Also the system allows robust to the light, as shown in the fourth image in Fig. 82.1 and the second image in Fig. 82.2.

Through the analysis above, we can see the system not only allows robust to light, rotation, and occlusion but also has a high rate of detection.

**Table 82.1** Collection methods of color images

Collection parameters	Parameter values
Posture	Frontal faces, partly rotated faces
Occlusion	Mustache, glass, hat
Illmination	Day mode, night mode
Backgrounds	Simple backgrounds, complex backgrounds

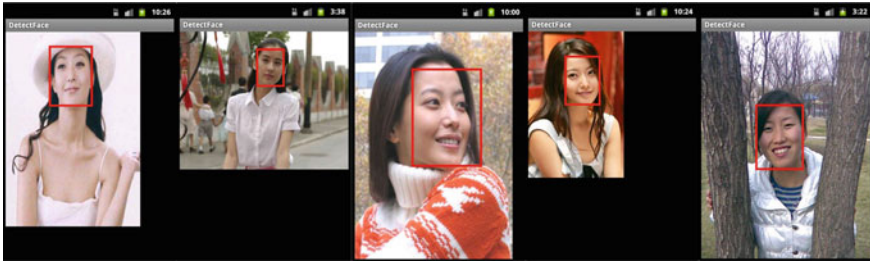


Fig. 82.1 Detection results of single-face test set



Fig. 82.2 Detection results of multiple faces test set

## 82.4 Conclusion

We have proposed an effective algorithm in this paper. For the purpose of segmenting skin regions, we suggest a method of cluster analysis-based segmentation. For measuring the similarity, we propose a method of template matching based weighted Euclid distance. The experimental results show the system posses robust to postures, occlusion, and illuminations, and can accurately detect and locate one or more faces. In the future we will make further improvement to optimize the system in real-time detection.

**Acknowledgments** The project was funded by Guangxi Key Laboratory of Spatial Information and Geomatics (No.1103108-27).

## References

1. Henry AR, Shumeet B, Takeo K (1998) Neural network-based face detection. *IEEE Trans Pattern Anal Mach Intell* 20(1):23–38
2. Liang LH, Ai HZ, He KZ, Zhang B (2001) Single rotated face localization based on affine template matching. *Chin J Softw* 12(1):94–102 (in Chinese)
3. Liang LH, Ai HZ, Xiao XP (2002) Face detection based on template matching and support vector machines. *Chin J Comput* 25(1):22–29 (in Chinese)

4. Krishnan N, Subban R, Babu CNK (2007) Skin detection using color pixel classification with application to face detection: a comparative study. In: Proceedings of the international conference on computational intelligence and multimedia applications
5. Jiang J, Zhang GL (2002) A knowledge-based approach for fast face detection. *Chin J Image Graph* 7(1):6–10 (in Chinese)
6. Yang M, Kriegman D, Ahuja N (2002) Detecting face in images: a survey. *Pattern Anal Mach Intell* 24(1):34–58
7. Vezhnevets V, Sazonov V, Andreeva A (2003) A survey on pixel-based skin color detection techniques. In: Proceedings of graphics
8. Song YZ, Zhang YY, Meng HD (2007) Research based on Euclid distance with weights of clustering method. *Comput Eng Appl* 43(4):179–180 (in Chinese)

# Chapter 83

## Image Mosaic Based on SURF and Results Optimization

Tie Jiang and Guibin Zhu

**Abstract** In order to improve the insufficiency of SURF algorithm, we present an auto-adjusted algorithm of image size based on phase-correlation. We detect the zoom relationship and translation co-efficiency between the images and modulate the unregistered image's scale to the same level as the original image. We obtain the Region of Interest (ROI) according to the translation parameter and then pre-treat the images. We propose an image matching method based on the saliency map. We calculate the saliency map in the Region of Interest and mark the interest points in the area by using SURF algorithm. In the part of image fusion, the method of gradated in-and-out which we commonly use is improved to eliminate seam. As the image exposure differences, we propose two methods to adjust the exposure of spliced image. One is based on mean value and the other is based on mean–variance specification. Through the experiment, the validity of the proposed methods is demonstrated and the quality of the mosaic image is better.

**Keywords** Image mosaic · Saliency map · SURF algorithm · Mean value · Mean–variance specification

### 83.1 Introduction

The automatic construction of large, high-resolution image mosaics is an active area of research in the fields of photogrammetry, computer vision, image processing, and computer graphics. Image mosaics can be used for many different applications [1–4]. It can be mainly divided into two categories [5]: feature-based methods, and featureless methods. Feature-based methods assume that feature

---

T. Jiang (✉) · G. Zhu  
Chongqing Key Laboratory of Emergency Communication, 13 team,  
Chongqing Communication Institute, Chongqing 400035, China  
e-mail: jiangtie.nupt@gmail.com

correspondences between image pairs are available, and utilize these correspondences to find transforms which register the image pairs. A major difficulty of these methods is the acquisition and tracking of image features.

Speeded Up Robust Features (SURF) algorithm [6] based on integral image was put forward by Herbert Bay, which approximated Laplacian-Gaussian operator by square filters based on integral image and constructed a Fast-Hessian matrix. The SURF algorithm improved the detection speed, but its performance was equal to SIFT algorithm.

In the paper, in order to reduce the running time of the SURF algorithm, before we use SURF algorithm for detecting feature points, we use a phase correlation method to calculate the overlap region of original image. We proposed one kind of feature detection method based on the saliency map. We calculate the saliency map in the Region of Interest and mark the interest points in the area by using SURF algorithm. We get fewer, more representative features. In the part of image fusion, the method of gradated in-and-out which we commonly use is improved to eliminate seam. As the image exposure differences, we propose two methods to adjust the exposure of spliced image. One is based on mean value and the other is based on mean-variance specification [7]. Experiments show that the proposed algorithm has achieved a well result.

## 83.2 Pre-Processing

### 83.2.1 Phase Correlation Algorithm

Phase correlation algorithm uses the cross-power spectrum to registration images and is used to get the translation factor initially. Imagine there are two images  $I_1$  and  $I_2$ , and the translation between them is as following:

$$I_2(x, y) = I_1(x - x_0, y - y_0) \quad (83.1)$$

The Fourier transformation:

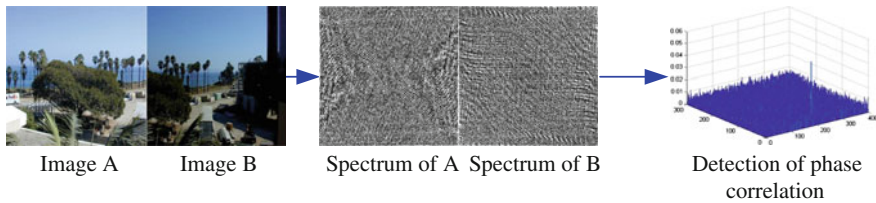
$$F_2(u, v) = F_1(u, v) \cdot e^{-j(ux_0 + vy_0)} \quad (83.2)$$

$F_1$  and  $F_2$  are the Fourier transformation of  $I_1$  and  $I_2$ . The cross-power spectrum is:

$$\frac{F_1^*(u, v)F_2(u, v)}{|F_1^*(u, v)F_2(u, v)|} = e^{-j(ux_0 + vy_0)} \quad (83.3)$$

$F_1^*(u, v)$  is the conjugate function of  $F_1(u, v)$ . We can get an impulse function  $\delta(x - x_0)(y - y_0)$  about the value of translation invariant  $x_0$  and  $y_0$  by using Fourier inverse transforming to formula (83.3). Figure 83.1 is an experiment to get the range shift of by using phase correlation algorithm. The preset value between





**Fig. 83.1** Use phase-correlation to estimate the offset parameters

image A and B is (188, 12). There is an obvious impulse at the range shift coordination compared to the flat site on the inverse conversion curve. So we can get the shift value via the location of the maximum value.

### 83.2.2 Saliency Map

The saliency map which Itti [8] proposed is a 2D image that shows the significance of each point of the image. The point of saliency map and the input image pixel have topology relationship. Different visual characteristics (such as color, orientation, etc.) have different contribution for the saliency map. But there is no interaction between them. The original image is decomposed by linear filtering for a variety of visual characteristics (brightness, red, blue, yellow, green, direction, etc.) and the respective characteristic diagram is decomposed into 9-storey Adams pyramid. Then characteristic maps in different scale and corresponding pixel subtract. We can see the contrast of the features and background. If the contrast is stronger, the significance of the point is greater. In order to eliminate the influence of the size of the neighborhood on the contrast, the contrast is calculated in multiple scales to generate a multi-scale feature maps. They use a function on the different characteristics of the image and filter them. Then they get the summation of response of various features, as the significant value of the point in the image. Dynamic neural networks were used for maximum point of significant value as the point of regard. At last, we obtain the saliency map. Figure 83.2 is a block diagram of Itti algorithm.

On the basis of the overlap region in the original images, we calculate the saliency map of the overlap region, then binaryzation image is obtained through adaptive threshold Segmentation. The image of the overlap region is matted with the binaryzation image, then we obtain a significant area, as shown in Fig. 83.3.

### 83.3 SURF Algorithm

SURF for interest point detection uses a very basic Hessian-matrix approximation. This lends itself to the use of integral images as made popular by Viola and Jones,

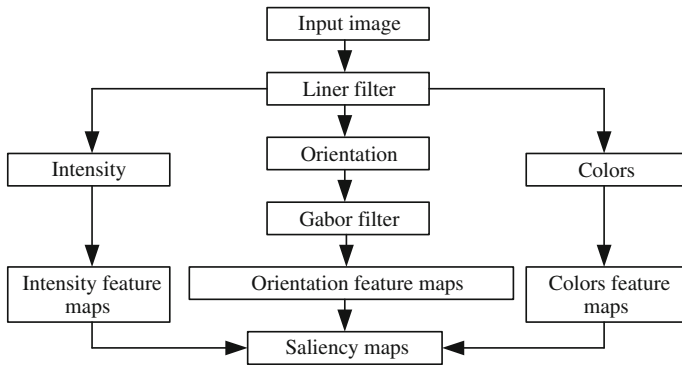


Fig. 83.2 Diagram of Itti model

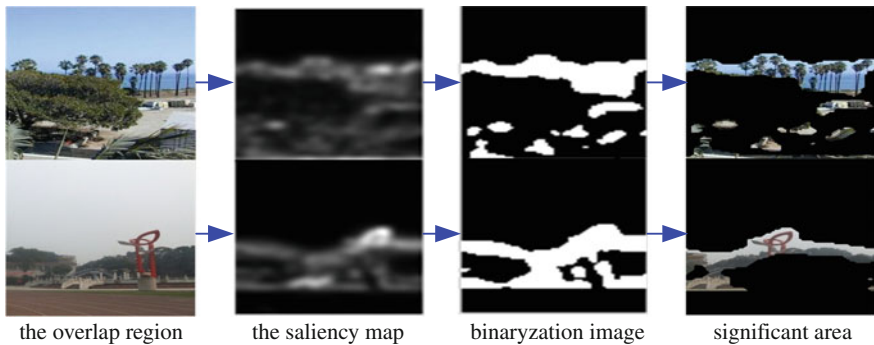
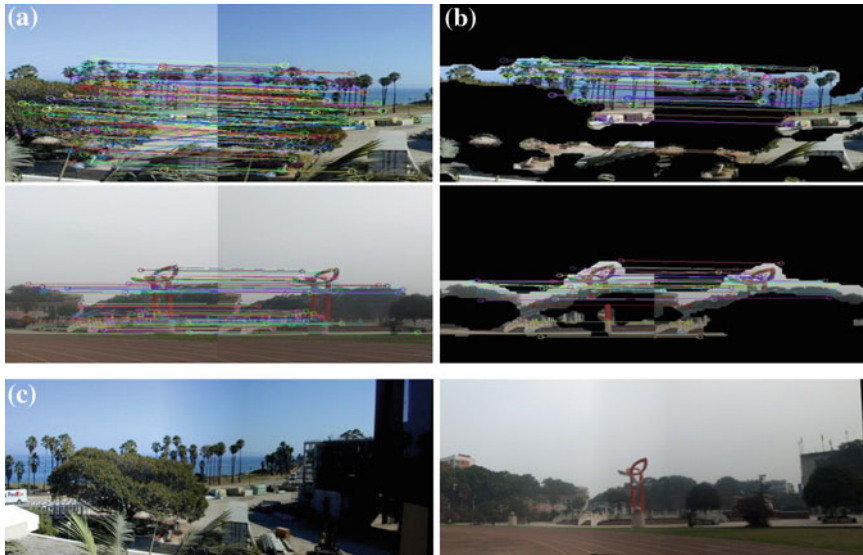


Fig. 83.3 The process of getting significant area of original Image

which reduces the computation time drastically. Integral images fit in the more general framework of boxlets, as proposed by Simard [9].

Comparing with SIFT, SURF algorithms employ slightly different ways of detecting features. SIFT builds an image pyramids, filtering each layer with Gaussians while increasing their sigma values and taking the difference. On the other hand, SURF creates a “stack” without 2:1 down sampling for higher levels in the pyramid, resulting in images of the same resolution [10]. Due to the use of integral images, SURF filters the stack using a box filter approximation of second-order Gaussian partial derivatives.

Integral images allow the computation of rectangular box filters in near constant time [10]. This paper used KNN to find the nearest neighbor with setting  $k$  to 2. RANSAC [11] is used to estimate a model for finding the minimize error matches set, which can maintain the correct matches by comparing the distance of the closest neighbor to that of second-closest neighbor [12]. If it is less than the distance ratio, then this paper maintains them, else removes. This paper decided to choose 0.5 as the distance ratio according to Lowe’s experiment in SIFT. This



**Fig. 83.4** The matches mosaic and results. **a** Matches based on the overlap region. **b** Matches based on the significant area. **c** Mosaic results

makes the repeatability larger, which means improving of the matching performance. More details of this algorithm can be seen on [13, 14].

The matches results and mosaic based on SURF features is shown in Fig. 83.4. Matches based on the overlap region are still some mismatch. Matches based on the significant area eliminate the mismatch and the feature points are reduced. The feature points are high precision and more reliable.

From Table 83.1, we see the contrast of time-consuming between SURF and method used in this paper. The speed of image mosaic based on the significant area is more than twice as high than the original SURF algorithm. The explanation is as following: Although we added the amount of work for phase section and saliency map, we delimit the ROI. And along with the preprocessing it decreased the workload dramatically. As a result the time-consuming was decreasing instead of increasing. This superiority can be performed better when we conduct the mosaic work aiming at big scenes and panoramic images.

## 83.4 Image Fusion

### 83.4.1 Weighted Function

Once we have found the best transformation  $H$ , we can blend the input image together with the reference image. To reduce visible artifact—that is, to hide the edges of the component images—we use a weighted average [15] with pixels near

**Table 83.1** Comparison of the different image mosaic method

Mosaic method	Original SURF	Matches based on the overlap region	Matches based on the significant area
Time (/s)	4.894	2.544	1.876
	4.687	2.484	2.125
	2.922	1.812	1.786

the center of each image contributing more to the final composite. The weighting function is:

$$f(x, y) = \begin{cases} f_1(x, y) & (x, y) \in f_1 \\ df_1(x, y) + (1 - d)f_2(x, y) & (x, y) \in (f_1 \cap f_2) \\ f_2(x, y) & (x, y) \in f_2 \end{cases} \quad (83.4)$$

where  $f_1, f_2$  is the reference image and input image,  $f$  is the mosaic image.  $d$  is the weighted value, and  $d = (x_2 - x)/(x_2 - x_1), x_1 \leq x \leq x_2, x_1, x_2$  is the minimum and maximum of  $x$  axis in the overlap region.

But the human eye is nonlinear on the response of luminance. The weighted factor  $d$  is linear. When the exposure of original image is very different, it is not very ideal that use the gradated in-and-out method to hide the edges of the component images. From Fig. 83.5c, we can see that the exposure of mosaic image is inconsistency. So we propose a trigonometric function to improve the weighting factor. That is a subsection weighting function:

$$d = \begin{cases} 1 - 0.5 \sin\left(\frac{x-x_1}{x_2-x_1}\right)\pi & x_1 \leq x < x_1 + \frac{x_2-x_1}{4} \\ \frac{(\sqrt{2}-2)(x-x_1)}{x_2-x_1} + \frac{3-\sqrt{2}}{2} & x_1 + \frac{x_2-x_1}{4} \leq x \leq x_1 + \frac{3(x_2-x_1)}{4} \\ 0.5 \sin\left(\frac{x-x_1}{x_2-x_1}\right)\pi & x_1 + \frac{3(x_2-x_1)}{4} < x \leq x_2 \end{cases} \quad (83.5)$$

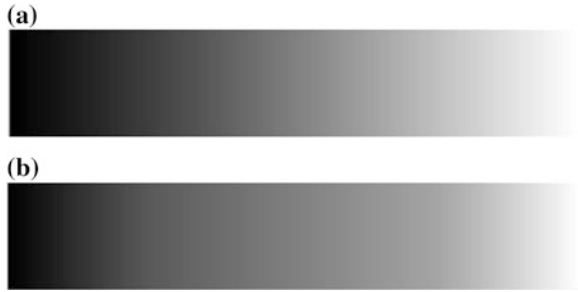
We use the improved weighted function to do simulation experiment. The experimental results are shown in Fig. 83.5.

In order to prove the optimization effect of new algorithm, we set two black and white image to fuse and compare the results. It fully reflects the gradient effect. Figure 83.6a shows the result that it uses the linear weighted function to fuse and Fig. 83.6b with improved subsection weighting function. We can see that the fusion image transition zone is small with the original function and the gray changes slowly in two sides. The middle part of the mosaic image changes quickly.



**Fig. 83.5** Weighting factor use subsection weighting function

**Fig. 83.6** Two different weighted function comparison. **a** Original weighted function. **b** Improved subsection weighting function



The image transition zone is bigger after using the improved subsection weighted function. The gray of the image can have the uniform changes in the entire interval and the transition is smoother. The mosaic image is more natural visual effect.

### 83.4.2 Exposure Adjustment

The result is still not very good from Fig. 83.6. The image in the overlapping region will still have brightness jump. It is not true for human eye. In order to make the mosaic image more smoother, we use Mean value method and Mean–variance specification method to adjust the exposure of the mosaic image. The mosaic image will be consistency of exposure, not obviously bright or dark.

#### 83.4.2.1 Mean Value Method

First, we calculate the mean values of the overlap region of two images and compare it, then calculate the D-value of them. We use the D-value to adjust the exposure of the mosaic image. When the image exposure is strong, we reduce it. On the contrary, we enhance it. Then we obtain a balanced effect of exposure.  $I_1, I_2$  is the two image,  $I_1$  is left and  $I_2$  is right. Their overlap region are  $M_1$  and  $M_2$ . The mean values of overlap pixels are  $\overline{M}_1$  and  $\overline{M}_2$ , then the public mean value is  $M = (\overline{M}_1 + \overline{M}_2)/2$ . The image mosaic is  $I'_1$  and  $I'_2$  after adjusting.

$$\begin{cases} I'_1 = I_1 - (\overline{M}_1 - M) \\ I'_2 = I_2 - (\overline{M}_2 - M) \end{cases} \tag{83.6}$$

#### 83.4.2.2 Mean–Variance Specification Method

Another method is Mean–variance specification. It is to use the average and variance of image. The mean and variance statistical parameters are commonly

used for color transmission. So we use of the two parameters to adjust the image mosaic. Statistical characteristics of one image is assigned to the other one. At last, it is consistency of the overall exposure of image mosaic. It can be expressed as:

$$I = \iint [\alpha I_2(x, y) + \beta - I_1(x, y)]^2 dx dy \tag{83.7}$$

where  $I_2$  is the image to be processed and  $I_1$  is the reference image,  $\alpha, \beta$  is overall linear transform coefficients, the formula is  $\alpha = \frac{\delta_{I_1 I_2}^2}{\delta_{I_2}^2}, \beta = M_{I_1} - \alpha M_{I_2}, \delta_{I_1 I_2}^2$  is two images for the covariance and  $\delta_{I_2}^2$  is image variance,  $M_{I_1}, M_{I_2}$  are the mean of the two images. The formula are as follows respectively:

$$\delta_{I_1 I_2}^2 = \frac{1}{S} \iint [I_1(x, y) - M_1][I_2(x, y) - M_2] dx dy \tag{83.8}$$

$$\delta_{I_2}^2 = \frac{1}{S} \iint [I_2(x, y) - M_2]^2 dx dy \tag{83.9}$$

$$M_1 = \frac{1}{S} \iint I_1(x, y) dx dy \tag{83.10}$$

$$M_2 = \frac{1}{S} \iint I_2(x, y) dx dy \tag{83.11}$$

The final results were:

$$I_2(x, y) = \alpha I_2(x, y) + \beta \tag{83.12}$$

The experimental results as shown in Fig. 83.7. The differences of mosaic image exposure are well adjusted by using the two method. The quality of mosaic image is further improved. It is good to meet the visual characteristics of human eyes.



**Fig. 83.7** The comparison of different method for exposure adjustment. **a** Exposure adjustment by mean value method. **b** Exposure adjustment by mean-variance specification method

## 83.5 Conclusions

In this paper we have presented an approach for image mosaic based on phase-correlation and saliency map. First the scaling and translation relationship is gained according to the correlation known by phase-correlation. Second significant area of the overlap region is obtained based on saliency map. Then the feature points are detected and matched just in this area, based on the SURF algorithm. The setting of ROI and adoption of preprocessing avoid the useless extraction and registration thus leading to additional speed-ups and improvement of the precision.

When the original image exposure differences, we proposed two methods to adjust it. Through adjusting exposure, the brightness jump is eliminated and the quality of the mosaic image is improved. The results are more in line with the human visual characteristics.

**Acknowledgments** This work is supported by the national natural science funds under Grant No. 61272043 and the open subject of Chongqing Key Laboratory of Emergency Communication under Grant 2012.

## References

1. Liu J-X, Chen Y-S, Chen L-F (2010) Fast and accurate registration techniques for affine and nonrigid alignment of MR brain images. *Ann Biomed Eng* 38:138–157
2. Brunet F, Gay-Bellile V, Bartoli A et al (2011) Feature-driven direct non-rigid image registration. *Int J Comput Vision* 93:33–52
3. Zhao H, Chen H, Yu H (2007) An improved fully-automatic image mosaic algorithm. *J Imag Graph* 12:336–342 (In Chinese)
4. Zheng S, Zhou Y (2009) A novel mosaic method for SAR image sequences with deficiency of overlap portion. *J Imag Graph* 14:2054–2060 (In Chinese)
5. Costas B (2001) A novel optical imaging method for the early detection, quantitative grading and mapping of cancerous and precancerous lesions of cervix. *IEEE Trans Biomed Eng* 48(1):96–105
6. Bay H, Tuytelaars T, Van Gool L (2006) SURF: Speeded Up Robust Feature. In: European conference on computer vision, 2006, pp 404–417
7. Papadopoulos A, Fotiadis DI (2008) Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques. *Comput Biol Med* 38(10):1045–1055
8. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans Pattern Anal Mach Intell* 20(11):1254–1259
9. Simard P, Bottou L, Haffner P, LeCun Y (1998) Boxlets: a fast convolution algorithm for signal processing and neural networks. In: NIPS
10. Bay H, Ess A, Tuytelaars T, Van Gool L (2008) SURF: Speeded Up Robust Features. *Comput Vis Imag Underst (CVIU)* 110(3):346–359
11. Fischler M, Bolles R (1981) Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun ACM* 24:381–395
12. Lowe D (2004) Distinctive image features from scale-invariant keypoints. *IJCV* 60(2):91–110
13. Zhang W, Kosecka J (2006) Generalized ransac framework for relaxed correspondence problems. In: International symposium on 3D data processing, visualization and transmission
14. Szeliski R (1996) Video mosaics for virtual environments. *IEEE Comput Graphics Appl* 16(2):22–30

# Chapter 84

## Improved Blind Source Separation Based on Non-Holonomic Natural Gradient Algorithm with Variable Step Size

Ce Ji, Baocheng Tang, Kun Yang and Mingbo Sha

**Abstract** The traditional natural gradient algorithm works badly when the source signal amplitude changes rapidly or becomes zero at a certain time. In addition, it cannot resolve very well the contradiction between the convergence speed and the error in steady state because the step-size is fixed. In order to solve the above problems, this paper proposes an improved blind source separation algorithm based on non-holonomic natural gradient by choosing an adaptive step-size and a suitable nonlinear activation function. Simulation result demonstrates that the new algorithm performance is superior to the traditional natural gradient algorithm.

**Keywords** Blind source separation · Non-holonomic natural gradient · Nonlinear activation function · Adaptive step size · Natural speech signals

### 84.1 Introduction

Blind source separation is one of the common research topics in signal processing and neural network field. It has been widely used in the filed of biomedical science [1], speech recognition and image processing [2, 3], array signal processing [4], Geophysical signal processing [5], financial forecasts [6] and so on. The natural gradient algorithm [7] that was first put forward in 1994 by Cichoki [8] is a very

---

C. Ji · K. Yang (✉)

College of Information Science and Engineering,  
Northeastern University, Shenyang, 110819 Liaoning, China  
e-mail: 1999yangkun@163.com

B. Tang

Troops 65521 PLA, Liaoyang, 111000 Liaoning, China

M. Sha

Allwin Telecommunication Co., Ltd, Shenyang, 110007 Liaoning, China



important approach for blind source separation. Later, Amari et al. approved the effectiveness of natural gradient algorithm theoretically [9–11]. The constraint  $E\{f_i(y_i)y_i\} = 1$  was imposed to the estimation signals in natural gradient algorithm in order to solve the uncertainty of amplitude. The constraint condition will also force the separate matrix to change rapidly when the source signals are non-stationary or their amplitudes rapidly change with time. Aimed at this problem, one non-holonomic constraint was added to natural gradient algorithm [12]. This algorithm was called non-holonomic natural gradient algorithm (NNG). When the magnitudes of source signals rapidly change with time or become zero in a certain period time, this algorithm still works well. Algorithm performance also does not fluctuate greatly even when the number of source signals is overestimated. However, the fixed step size was adopted for the non-holonomic natural gradient algorithm [12]. The disadvantage of fixed step size is slow convergence speed. The signal separation is unsuccessful if the oversized step is chosen.

In order to solve the problem above, a new non-holonomic natural gradient algorithm with adaptive variable step and tangent non-linear active function is presented in this paper. The improved algorithm has prominent effect in natural speech signal separation.

## 84.2 Blind Source Separation and Non-Holonomic Natural Gradient Algorithm

### 84.2.1 Blind Source Separation

Blind source separation is the method of recovering the original signals from the multiple observed signals mixed randomly. The original signals are from different sources respectively, so usually they are mutually independent. Hypothesize that  $n$  statistical signals  $s_i(t)$  ( $i = 1, 2, \dots, n$ ) are linearly instantaneously mixed.  $n$  mixed signals (or observed signals)  $\mathbf{x}_i(\mathbf{t})$  are denoted as:

$$\mathbf{x}_i(\mathbf{t}) = \sum_{j=1}^n \mathbf{a}_{ij} \mathbf{s}_j(\mathbf{t}) \quad (84.1)$$

where  $\mathbf{a}_{ij}$  is mixing coefficient. In vector form, Eq. (84.1) is equal to:

$$\mathbf{x}(t) = \mathbf{A} \mathbf{s}(t) \quad (84.2)$$

In Eq. (84.2),  $\mathbf{s}(t) = [s_1(t), \dots, s_n(t)]^T$  is  $n \times 1$  unknown source signals vector,  $\mathbf{x}(t) = [x_1(t), \dots, x_n(t)]^T$  is  $n \times 1$  observed signals vector, and  $\mathbf{A}$  is  $n \times n$  unknown mixture matrix.

Blind source separation is the process of establishing the objective function of separating matrix  $\mathbf{W}$  and making output signal  $y(t)$  approach the original signal  $s(t)$  possibly based on the assumption of statistical independence of each components  $s_i(t)$  ( $i = 1, 2, \dots, n$ ). Therefore, the  $n \times n$  separating matrix  $\mathbf{W}$  is given by:

$$y(t) = \mathbf{W}x(t) = \mathbf{W}\mathbf{A}s(t) = \mathbf{G}s(t) \quad (84.3)$$

In Eq. (84.3),  $y(t)$  is known as estimated signal or separated signal. If  $\mathbf{W}$  which is obtained from some learning algorithm can make global matrix  $G = I$  ( $I$  is a  $n \times n$  unit matrix), then  $y(t) = s(t)$ , and the source signals are recovered.

### 84.2.2 Non-Holonomic Nature Gradient Algorithm

The classical approach to resolve the non-restraint optimization problem of a multivariate function is the steepest descent method, also called gradient descent algorithm. Amari proposed the non-holonomic nature gradient algorithm based on natural gradient algorithm. Actually, non-holonomic nature gradient uses a specific diagonal matrix  $\Lambda$  to replace identity matrix  $I$  in natural gradient algorithm. So, the non-holonomic nature gradient can be expressed as:

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k)[\Lambda - F(Y)Y^T]\mathbf{W}(k) \quad (84.4)$$

In Eq. (84.4),  $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_n\}$  is the diagonal positive definite matrix whose elements are  $\lambda_i = F(Y_i)Y_i^T$ .  $F(Y)$  is the activation function of gradient algorithm.  $\mu(k)$  is the step factor. Usually, we use nonlinear odd function as activation function according to the Gaussian character of separating signals, for example  $F(Y) = Y^3$  (sub-Gaussian) or  $F(Y) = \tanh(Y)$  (super-Gaussian).

From the Eq. (84.4), we can find that non-holonomic nature gradient algorithm still works well when the magnitudes of source signals rapidly change over time or when some of them become zero for a while. Even when the number of source signals is overestimated, algorithm performance also does not fluctuate greatly. However, the natural gradient algorithm does not have the skill and it will amplify the small components that do not exist in the source signals and cause the degradation of capabilities. Amari had proved this in detail [12]. The simulation results in [12] demonstrated that the separation effect and stability of non-holonomic nature gradient algorithm is better than natural gradient algorithm, so we use non-holonomic nature gradient algorithm to separate natural speech signals in this paper.

## 84.3 The Non-Holonomic Nature Gradient Algorithm Based on Variable Step

### 84.3.1 The Selection of Non-Linear Function

Non-linear function not only limits the step size, but also directly influences the algorithm's stability. In ideal conditions, the source function of source signal is chosen as non-linear function. However, the source function cannot be gotten in practical applications. Generally, when non-linear function and source function have little difference, the algorithm can reach convergence. In the literature [13], the active function was chosen as follows:

$$\varphi(y) = 2 \tanh(y) \quad (84.5)$$

In this paper, the tangent active function is chosen:

$$\varphi(y) = \begin{cases} 1 & y > \pi/4 \\ \frac{11}{4} \tan(y) & -\pi/4 \leq y \leq \pi/4 \\ -1 & y < -\pi/4 \end{cases} \quad (84.6)$$

The performances of two active functions will be compared in the following experiment of separating natural speech signal.

### 84.3.2 The Selection of the Controlling Parameters of Step Adaptation

In Eq. (84.4),  $\mu(k)$  is the step factor that affects the convergence speed and stability of the algorithm. In the adaptive algorithm of blind separation, step size plays a key role to the algorithm convergence. A better way is to adjust the step size with learning rate, which changes with time. So a harmony between convergence speed and steady performance is obtained. The purpose of any time-varying step size processing is to achieve the fastest convergence by increasing the step size to one large stable value. In fact, when these parameters get to the neighborhood of the best convergence point, the step size should decrease correspondingly in order to reduce maladjustment error.

The separating matrix  $\mathbf{W}$  can converge to a certain stable neighborhood, but it fails to converge to a fixed value. Therefore, the distance between separating matrix  $\mathbf{W}$  and the best separating matrix  $\mathbf{W}_{opt}$ , which is defined as separating degree, can be used to adjust step size [14]. Because  $\mathbf{W}_{opt}$  cannot be given beforehand, the similar ways to replace  $\mathbf{W}_{opt}$  is as follows:

$$\Delta \mathbf{W}(k) = \|\mathbf{W}(k+1) - \mathbf{W}(k)\|_F^2 \quad (84.7)$$

In Eq. (84.7),  $\|\bullet\|_F$  is Frobenius norm.

The aim of blind source separation is to search the optimal matrix  $\mathbf{W}_{opt}$ , which meets the next equation:

$$\mathbf{W}(k+1) = \mathbf{W}(k) = \mathbf{W}_{opt} \quad (84.8)$$

The expression of updating step size can be described as follows:

$$\mu(k+1) = \alpha(k)\mu(k) \quad (84.9)$$

The exponent function  $e^{-k}$  has a very good convergence performance. Therefore, increasing or decreasing  $\alpha(k)$  exponentially may show a very good convergent effect. The value of  $\alpha(k)$  can be chosen according to the following equation:

$$\alpha(k) = \begin{cases} 1 + \gamma e^{-k} & \Delta W(k) < \Delta W(k-1) \\ \frac{1}{1+\beta e^k} & \Delta W(k) > \Delta W(k-1) \\ 1 & else \end{cases} \quad (84.10)$$

In Eq. (84.10),  $0 < \beta < 1$ ,  $0 < \gamma < 1$ .  $\beta$  is used to control the stable error when converging.  $\gamma$  is used to control convergence speed. The new non-holonomic natural gradient algorithm with adaptive step size can be described as:

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \alpha(k)\mu(k)[\mathbf{A} - F(Y)Y^T]\mathbf{W}(k) \quad (84.11)$$

## 84.4 Analysis of simulation results

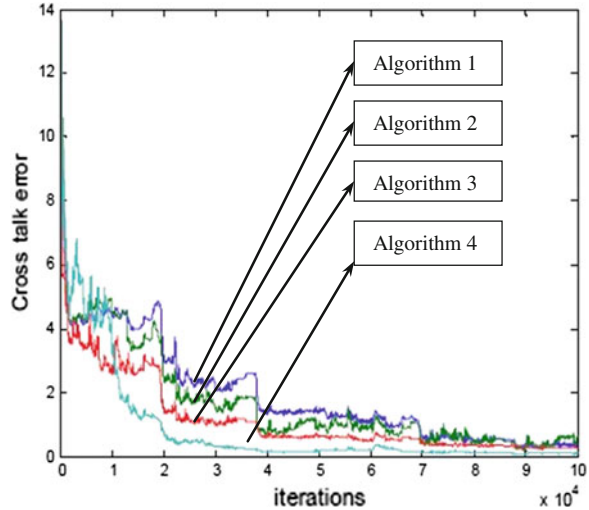
Suppose that there are 4 speech source signals. The first signal is the staccato alarm, the 2nd signal and 3rd signal are male announcer's voices, and the 4th signal is female announcer's voice. The duration of every signal is 10 s. Sampling rate is 10 kHz. There are  $10^5$  sample points. Firstly, the pretreatment of zero-mean and whitening are done to speech signals [15]. Let  $W(0) = 10I$ , and mixed matrix  $A$  is given by:

$$A = \begin{bmatrix} 0.71812 & 0.08774 & 0.85184 & 0.01423 \\ 0.56919 & 0.44348 & 0.75948 & 0.59618 \\ 0.46081 & 0.36630 & 0.94976 & 0.81621 \\ 0.44531 & 0.30253 & 0.55794 & 0.97709 \end{bmatrix}$$

In order to compare the separation effect of different algorithms, cross-talking error  $E$  [16] is adopted to measure the performance of the algorithm in this paper. It is defined as follows:

$$E = \sum_{i=1}^N \left[ \sum_{j=1}^N \frac{|P_{ij}|}{\max_k |P_{ik}|} - 1 \right] + \sum_{j=1}^N \left[ \sum_{i=1}^N \frac{|P_{ij}|}{\max_k |P_{kj}|} - 1 \right] \quad (84.12)$$

**Fig. 84.1** Cross talk error curve of four algorithms



Here  $\mathbf{P} = (P_{ij}) = \mathbf{W}\mathbf{A}$  is the performance matrix of the system.

Algorithm 1: using formula (84.5) as the active function, adopting the constant step size  $\mu(k) = 0.0006$ ;

Algorithm 2: using formula (84.6) as the active function, adopting the constant step size  $\mu(k) = 0.0006$ ;

Algorithm 3: using formula (84.5) as the active function, adopting the adaptive step size  $\mu(k+1) = \alpha(k)\mu(k)$ ,  $\beta = 0.02$ ,  $\gamma = 0.4$ ;

Algorithm 4: using formula (84.6) as the active function, adopting the adaptive step size  $\mu(k+1) = \alpha(k)\mu(k)$ ,  $\beta = 0.02$ ,  $\gamma = 0.4$  (Fig. 84.1).

From the point of actual effect, four algorithms can separate speech signals. The following conclusion can be obtained by means of specific analysis:

- Whether using fixed step size or adaptive step interval, the improving non-holonomic natural gradient algorithm based on tangent activation function is perfectly suitable to blind source separation of speech signals in this paper. By further analyzing the performance curve, we can see that Algorithm 2 is better than algorithm 1 before  $6 \times 10^4$  iterations, and Algorithm 4 is better than Algorithm 3 in the premise of the same step size.
- Comparing with the non-holonomic natural gradient algorithm with fixed step size, the separation properties of non-holonomic natural gradient algorithm with adaptive step size is greatly enhanced no matter which kind of nonlinear activation function is used. Therefore, the adaptive step size of this paper is effective to the non-holonomic natural gradient algorithm.

## 84.5 Conclusion

This paper studies the blind source separation of non-holonomic natural gradient algorithm based on tangent activation function. On the basis, a new non-holonomic natural gradient algorithm with adaptive step is proposed. Finally, simulation results show the improved algorithm has more desirable convergence rate and steady performance.

**Acknowledgments** 1. National Natural Science Foundation of China (11273001, 61074073, 61273164).

2. New Century Excellent Talents in University (NCET-10-0306), Ministry of Education, China.

## References

1. Hesse CW, James CJ (2006) On semi-blind source separation using spatial constraints with applications in EEG analysis. *IEEE Trans Biomed Eng* 53(12):2525–2534
2. Rivet B, Girin L, Jutten C (2007) Mixing audiovisual speech processing and blind source separation for the extraction of speech signals from convolutive mixtures. *IEEE Trans Audio Speech Lang Process* 15(1):96–108
3. Abolghasemi V, Ferdowsi S, Sanei S (2012) Blind separation of image sources via adaptive dictionary learning. *IEEE Trans Image Process* 21(6):2921–2930
4. Duarte LT, Jutten C, Moussaoui S (2009) A Bayesian nonlinear source separation method for smart ion-selective electrode arrays. *IEEE Sens J* 9(12):1763–1771
5. Wei W (2010). Seismic data noise—suppression technique based on blind source separation. *Comput Simul* 27(11):274–277 (in Chinese)
6. Duan Z, Ma Y, Peng B et al (2009) Computer simulation analysis of commercial sales based on blind source separation. *Comput Simul* 26(11):308–311 (in Chinese)
7. Ji C, Yu P, Yu Y (2010) Blind source separation based on improved natural gradient algorithm. *Proc Eighth World Congr Intell Control Autom* 31(8):6904–6807 (in Chinese)
8. Cichocki A, Unbehauen R, Moczczynski R, Rummaert E (1994) A new on-line adaptive learning algorithm for blind separation of source signals. In: *Proc. ISANN 94*, pp 406–411
9. Amari S (1998) Natural gradient works efficiently in learning. *Neural Comput* 10:251–276
10. Amari S, Cichocki A (1998) Adaptive blind signal processing. *Neural network approaches*. *Proc. IEEE* 86:2026–2048
11. Yang HH, Amari S (1997) Adaptive on-line learning algorithms for blind separation: maximum entropy and minimum mutual information. *Neural Comput* 7:1457–1482
12. Amari S, Chen TP, Cichocki A (2000) Non-holonomic constraints in learning algorithms for blind source separation. *Neural Comput* 12:1463–1484
13. Jacob B (2000). *An introduction to blind source separation of speech signals*. Acoustic signal processing for telecommunication. Kluwer Academic, Boston, pp 321–329
14. Ji C, Tang B, Zhu L (2011) Blind source separation of speech signals based on variable step length natural gradient algorithm. *J Central South Univ (Nat Sci)* 42(1):661–665 (in Chinese)
15. Pei X (2007) Blind source separation based on variable step length natural gradient algorithm. *Warship Electron Counter* 30(4):65–68 (in Chinese)
16. Yuan L, Wang W, Chambers JA (2005) Variable step-size sign natural gradient algorithm for sequential blind source separation. *IEEE Signal Process Lett* 12(8):589–592

# Chapter 85

## Multi-Level Fingerprint Classification Based on Average Frequency of Ridges for Large Scale Fingerprint Database

Xiaoqi Peng and Yunfei Zhong

**Abstract** In order to improve the recognition speed, accuracy and robustness of the Automatic Fingerprint Identification System based on large scale fingerprint database, a multi-level fingerprint classification method was proposed based on three independent classification features, in which the quality evaluation indexes algorithm of fingerprint images was introduced to evaluate the quality of input fingerprints. For those good quality fingerprints, we classify them into three features according to fingerprint pattern type, ridge count between singular points and average frequency of ridges in central region, respectively, which could decrease the retrieval space gradually. Experimental results on NIST DB4 show that the proposed classification algorithms with redundancy mechanism have high retrieval efficiency and strong robustness, which provides a rapid and effective index approach for the large scale fingerprint database.

**Keywords** Multi-level classification · Fingerprint quality evaluation index · Ridge count · Average frequency

---

X. Peng · Y. Zhong (✉)

School of Information Science and Engineering, Central South University,  
No. 932 LuShan Road, Changsha City, China  
e-mail: maczone@163.com

X. Peng

e-mail: pengxq126@126.com

Y. Zhong

School of Packaging and Materials Engineering, Hunan University of Technology,  
No. 88 TaiShan West Road, Zhuzhou City, China

## 85.1 Introduction

Biometrics (or biometric authentication) refers to the identification of humans by their characteristics or traits, which is often used in computer science as a form of identification and access control with the rapid development of computer networks and Internet technology [1]. A number of biometric traits are applied in various applications. Each trait has its own strengths and weaknesses and the choice typically depends on the application. No single trait is expected to effectively meet the requirements of all the applications, but fingerprint has a nice balance among all the desirable properties, and fingerprint recognition is one of the most mature biometric technologies and is suitable for a large number of recognition applications because of high distinctiveness, permanence and performance [2].

It is especially difficult in a large-scale identification system in which the number of enrolled users is huge (e.g., in the millions). Large volumes of fingerprints are collected and stored every day in a wide range of applications, particularly in forensics and government applications, e.g., background check of employees. Automatic identification based on fingerprints requires matching the input (query or test) fingerprint with a large number of templates stored in a database. To reduce the search time and computational complexity, it is desirable to classify these fingerprints in an accurate and consistent manner. Fingerprint classification refers to the problem of assigning a fingerprint to a class in a consistent and reliable way. Although fingerprint matching is usually performed according to local features (e.g., minutiae), fingerprint classification is generally based on global features, such as global ridge structure and singularities. Because common fingerprint classification characteristics exist in smaller between-class variance and larger within-class variance, low quality fingerprint and other reasons, automatic fingerprint classification is always a difficult task in the field of pattern recognition [3].

Fingerprint classification problem has attracted a significant amount of interest in the scientific community due to its importance and intrinsic difficulty, and a large number of papers have been published on this topic during the last 30 years. The fingerprint classification literature is surveyed in detail and the proposed methods are categorized into one or more of the following families: rule-based approaches, syntactic approaches, structural approaches, statistical approaches, neural networks-based approaches, and multi-classifier approaches [4]. Most of these classification algorithms built on the basis of Henry's 5 classification model or its variants, whose classification quantity is less. When the fingerprint database is large, these algorithms may not be effective in reducing the number of fingerprints to be matched.

In order to improve the efficiency of the Automated Fingerprint Identification System for large scale fingerprint database, a multi-level fingerprint classification algorithm is proposed based on independent classification features. Quality evaluation indexes may be applied to evaluate the quality of input fingerprints, remind the users to re-enter fingerprints if the quality is poor. For the good quality



fingerprints, three independent features including fingerprint pattern type, number of ridges between singular points and average frequency of ridges in central region are separately used to realize multi-level classification, which could decrease the retrieval space gradually. Experimental results show that the proposed classification algorithm has high retrieval efficiency and strong robustness, which provides a rapid and effective index mechanism for the large-scale fingerprint database. It has good real-time performance.

## 85.2 Outline

Classification characteristics of fingerprint database should have good stability and separability in order to achieve multi-class classification index. In this paper fingerprint common class of the Galton-Henry classification scheme, ridges count between neighboring singular points and average frequency of ridges in central region are selected as classification and retrieval features, which are inherent characteristics of the fingerprint itself. The steps of fingerprint classification and retrieval scheme are given below.

- Assuming the entire fingerprint database to retrieve is set A.
- Quality evaluation to every fingerprint image, the good for Galton-Henry classification is set B and the poor to be discarded.
- Ridge count calculating to get set C.
- Average frequency calculating to get set D from C.
- Detailed feature matching for set D.

Fingerprint image quality evaluation should be done at first because low quality fingerprint images seriously affect fingerprint recognition accuracy. The combination of comprehensive evaluation methods with macroscopic and microscopic characteristics were employed to determine the eligibility of the fingerprint image quality as described in literature [5]. Complex filtering method can locate the core point of the fingerprint image and triangular point [6] and provide other information of singular point. The fingerprint images can be divided into six classes according this information [7] including arch, tented arch, left loop, right loop, whorl and a miscellaneous type which means this type cannot belong to anyone of the five classes.

Ridge count has been typically used in forensic matching because of the difficulty of human experts to work in the Euclidean space [8]. When the core point and a triangular point can be precisely positioned, the ridge count is a fixed value which is against translation, rotation, etc. deformation factors. So ridge count can be used as classification feature discrimination with robustness for loop and whorl fingerprints.

### 85.3 Average Frequency of Ridges Methods

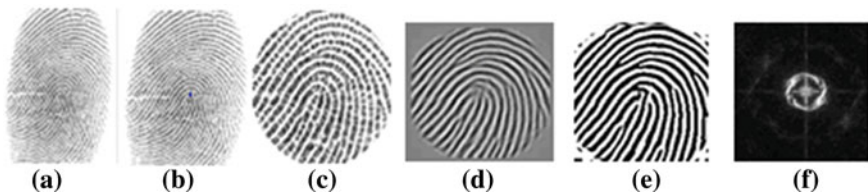
A fingerprint is the reproduction of the exterior appearance of the fingertip epidermis. The most evident structural characteristic of a fingerprint is a pattern of interleaved ridges and valleys. Different fingerprint has different average frequency, therefore the frequency of ridges can be used as a fingerprint classification feature.

#### 85.3.1 Spectrum Analysis and Preprocessing of Fingerprint Images

In general nonlinear deformation of the fingerprint center region is usually small with the sharpness of the ridge. To improve the accuracy of the calculation of the average frequency, reduce the effects of poor quality of the fingerprint image and the computational complexity, we selected only the circular center area to deal with spectrum analysis (the centre of a circle is the fingerprint center point, radius is the distance between the center point and the edge of the fingerprint image pixels.). The circular area intercepted was enhanced in order to reduce the impact on the calculation of image's gray level, and then which was binarized. The binarization area was performed by Fourier transform to obtain the average frequency of the ridge in the frequency domain of the fingerprint image center region by the spectrum analysis. The above processing was shown in Fig. 85.1.

#### 85.3.2 Frequency Spectrum Characteristics Analysis and Average Frequency Calculating of Ridges

Let  $f(x,y)$  represent the gray value of the  $N \times N$  digital image, where  $(x,y)$ ,  $F(u,v)$  is two-dimensional Discrete Fourier Transform of  $f(x,y)$ , then:



**Fig. 85.1** Spectrum analysis of fingerprint center region (from left to right: **a** original fingerprint image; **b** locating the center point; **c** clipped image by circle; **d** enhanced image; **e** binarized image; **f** frequency spectrum image)

$$F(u, v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) e^{-2\pi j[(ux+vy)/N]} \tag{85.1}$$

where  $u, v \in \{0, \dots, N - 1\}$ .

Fingerprint center region was operated by the two-dimensional Discrete Fourier Transform to obtain centrosymmetric spectrum which was shown in Fig. 85.1f. We set the center point at the origin, the distance between the origin and the destination point be the radius  $r$ , the angle of the target point be  $\theta$  and the frequency spectrum in the polar coordinate system be  $F(r, \theta)$ . The average frequency was calculated as follows:

- Spectrum amplitude distribution probability density function  $p(r, \theta)$

$$p(r, \theta) = \frac{|F(r, \theta)|^2}{\int_0^\infty \int_0^{2\pi} |F(r, \theta)|^2 drd\theta} \tag{85.2}$$

where  $|F(r, \theta)|^2$  is the energy value of target point in frequency domain, and  $\int_0^\infty \int_0^{2\pi} |F(r, \theta)|^2 drd\theta$  is the whole energy value.

- Spectrum amplitude probability density function  $p(r)$  of the edge

$$p(r) = \int_0^{2\pi} p(r, \theta) d\theta, \quad r \in \left[0, \frac{NFFT}{2} - 1\right] \tag{85.3}$$

$p(r)$  is the spectral energy distribution density on the ring with radius  $r$ .

- Average frequency

$$\bar{r} = \int_{10}^{90} \frac{p(r)}{p_{-all}} r dr \tag{85.4}$$

where  $p_{-all}$  is the spectral energy distribution density on the ring with radius  $r \in [10, 90]$ . The spectral structure analysis and experiments show that the highest stability of the average frequency of the calculation result can be obtained when the integration interval is taken in the range of  $[10, 90]$ .

### 85.3.3 Result Analysis

Nine different fingerprint images corresponding to three fingers were randomly selected from FVC2002 fingerprint database. Follow the above steps to spectrum analysis and the corresponding average frequency were obtained as shown in Table 85.1.

**Table 85.1** The average frequency of different fingerprint images from the same finger

Finger no.	Images	$\bar{r}$
13	1	38.848
13	2	38.495
13	3	38.173
14	1	33.298
14	2	33.390
14	3	32.468
15	1	35.531
15	2	35.179
15	3	35.238

The average frequency has only small difference for the different fingerprint images from the same finger undergoing rotation, translation or different pressure. There is a big difference for the different fingerprint images from the different finger. These indicate that the average frequency of the fingerprint ridge can be used as a stable, distinguishable classification feature.

If there is quality difference between the same source fingerprints, it may lead that the calculation of the average frequency is not exactly. In this paper, we proposed a redundancy classification strategy which applies the average frequency to establish three categories; thereby the number of fingerprints to be matched was further reduced. The fingerprints without being classified in the second classification could be also classified by this way, such as arch, tented arch and the miscellaneous type. Because any fingerprint can calculate the average frequency of the ridges, it can be suited for any fingerprint classifications. Search can be carried out in the same category of the average frequency of the ridges. If the search failed, the second search can be carried out in a similar average frequency category.

In the multi-level classification system, classification error has cascading properties. Because fingerprint pattern type categories and ridge count between the singular points have strong stability and consistency, the first and second level classification features were adopted respectively. Therefore, the average frequency of the ridges may be reasonably treated as the third level classification features because the stability of the average frequency of the ridges is weaker than that of the Galton-Henry classification scheme and ridges count between neighboring singular points.

## 85.4 Performance

NIST released two fingerprint databases well suited for development and testing of fingerprint classification systems: DB4 and DB14. Both databases consist of 8-bit grey-level images of rolled fingerprint impressions scanned from cards; two

different fingerprint instances (F and S) are presented for each finger. Each fingerprint was manually analyzed by a human expert and assigned to one of the five classes: Arch (A), Left loop (L), Right loop (R), Tented arch (TA), and Whorl (W). Actually, in DB4, some ambiguous fingerprints (about 17 %) have an additional reference to a “secondary” class. DB4 contains 2,000 fingerprint pairs which uniformly distributed in the five classes. NIST DB4 became standard benchmarks for fingerprint classification and most of the algorithms published in the last decade were tested by one of these databases. In this paper, we randomly selected 2000 fingerprint image for quality evaluation from the NIST DB4 and we found 214 fingerprint images is unqualified. The performance of the fingerprint classification system was tested for the remaining 1786 fingerprint images, and first level classification results are shown in Table 85.2.

The error rate on DB4 of the first classification level approach was 10.5 %, which can be calculated from Table 85.2. The accuracy of the proposed method shows a distinct advantage compared to 14.6 % mentioned in literature [9] and 15.7 % in [10].

1065 performance test images of the second level classification were selected from the remaining 1786 fingerprint including L, R and W. Because W-type fingerprint has more than two singular points, the experiment was to calculate the ridge count between the nearest singular points. There are errors during the calculation due to the effects of noise and the thinned process may cause part of the ridge distortions. The process of singular points position calculation may also produce errors too. The redundant classification strategy was put forward in order to improve the robustness and accuracy of classification algorithm. Let the maximum ridge count be  $n$ , the fingerprints of ridge count from 0 to 2 were divided into a class, referred to as  $R_1$ . The fingerprints of ridge count from 2 to 4 were divided into a class, labeled as  $R_2$ , and so on, until the fingerprints of ridge count from  $n - 2$  to  $n$  were divided into the last category. Thus the fingerprints were divided into 13 categories. The results are shown in Table 85.3.

1786 fingerprint images were classified by the third level classification method. The test result showed that the average frequency of ridges distributed in the range 17–49 and the 34–46 was about 74.2 %. The redundant classification strategy was also employed. The results are shown in Table 85.4.

The second and the third classification are continuous classification and the accuracy rate can be validated by matching experimental. If the retrieval is within

**Table 85.2** The classification results of the first level

True class	Hypothesized class					
	L	R	W	A	TA	Miscellaneous
L	335	3	9	13	0	22
R	3	313	8	10	2	15
W	8	6	350	2	3	9
A	2	0	2	424	3	2
TA	6	5	9	36	176	10

**Table 85.3** The statistics results of the second classification level

Class	Percentage (%)	Class	Percentage (%)
R <sub>1</sub>	12.86	R <sub>8</sub>	6.29
R <sub>2</sub>	11.64	R <sub>9</sub>	4.88
R <sub>3</sub>	12.30	R <sub>10</sub>	4.32
R <sub>4</sub>	14.27	R <sub>11</sub>	1.13
R <sub>5</sub>	12.58	R <sub>12</sub>	0.85
R <sub>6</sub>	11.46	R <sub>13</sub>	0.56
R <sub>7</sub>	6.86		

**Table 85.4** The statistics results of the third classification level

Class	Average frequency	Percentage (%)	Class	Average frequency	Percentage (%)
F <sub>1</sub>	17–22	4.4	F <sub>6</sub>	37–39	19.5
F <sub>2</sub>	22–26	3.6	F <sub>7</sub>	39–41	15.1
F <sub>3</sub>	26–30	5.5	F <sub>8</sub>	41–43	12.2
F <sub>4</sub>	30–34	9.6	F <sub>9</sub>	43–46	9.6
F <sub>5</sub>	34–37	17.8	F <sub>10</sub>	>46	2.7

the maximum search radius, the matching accuracy rate can reach 100 %. In practice fixed radius or incremental search may be used for indexing the database in a continuous classification technique.

In order to evaluate the performance of the classification algorithm mentioned above, we introduce the classification efficiency function [11]. If using  $k$  classification features for fingerprint classification and  $w_i$  class fingerprint probability is  $p(w_i)$ , the entropy of classified information:

$$H(w) = - \sum_{i=1}^k p(w_i) \log_2 p(w_i). \tag{85.5}$$

The classification efficiency function of classification algorithms was defined as

$$E = 2^{H(w)} = 2^{- \sum_{i=1}^k p(w_i) \log_2 p(w_i)} = 1 / \sum_{i=1}^k p^2(w_i). \tag{85.6}$$

Equation 85.6 shows that classification algorithms can divide the fingerprint database into  $E$  categories. The classification efficiency function from the first level to the third level is 5.13, 9.30 and 7.53. The classification efficiency of multi-class classification algorithm developed in this paper is  $E_1 \times E_2 \times E_3$  and would be 359.25, which means that classification system has the ability to divided fingerprints into 359.25 categories. Relative to the traditional five major categories of classification, the classification efficiency has improved significantly. This method

can meet the classification requirements of large scale fingerprint database and significantly improve the retrieval speed and matching accuracy.

## 85.5 Conclusion

The rapid recognition technology of the large scale fingerprint database is a hot and difficult field of research. In this paper, the multi-class classification algorithms based on three independent classification features were put forward in order to improve recognition efficiency. The experiments result shows that the classification algorithms are of good retrieval efficiency, robustness and small matching recognition space. It provides a fast and efficient retrieval and matching mechanism and has a strong practical for large scale fingerprint database.

**Acknowledgments** This work was supported by Natural Science Foundation of Hunan Province, (Grant No. 10JJ2048 and 12JJ9043), Science and Technology Project of Hunan Province (Grant No. 2012FJ3038), Training Project of Hunan Industrial Application of Higher Education (Grant No. 10CY006) and Natural Science Research Project of Hunan University of Technology (Grant No. 2011HZX03), China.

## References

1. Jain A, Hong L, Pankanti S (2000) Biometric identification. *Commun ACM* 43(2):91–98
2. Maltoni D, Maio D, Jain A, Prabhakar S (2009) Handbook of fingerprint recognition, 2nd edn. Springer, London
3. Ahmad F, Dzulkifli M (2009) A review on fingerprint classification techniques. In: International conference on computer technology and development, pp 411–415
4. Yager N, Admin A (2004) Fingerprint classification: a review. *Pattern Anal Appl* 7(1):77–93
5. Hu M, Li DC (2010) A method for fingerprint image quality estimation (in Chinese). *Comput Technol Dev* 20(2):125–128
6. Nilsson K, Bigun J (2003) Localization of corresponding points in fingerprints by complex filtering. *Pattern Recogn Lett* 24(13):2135–2144
7. Liu MH (2009) Fingerprint classification based on singularities techniques. In: International conference on computer technology and development, pp 1–5
8. Watson CI, Wilson CL (1992) NIST Special database 4, fingerprint database. U.S. National Institute of Standards and Technology
9. Karu K, Jain A (1996) Fingerprint classification. *Pattern Recogn* 29(3):389–404
10. Zhang QZ, Yan H (2004) Fingerprint classification based on extraction and analysis of singularities and pseudo ridges. *Pattern Recogn* 37(11):2233–2243
11. Shu Z (2006) The research of fingerprint classification based on genetic algorithm (in Chinese). Huazhong Univ Sci Technol

# Chapter 86

## Two Improved Edge Coloring Algorithms for Data Migration

Gangfeng Huang, Maishun Yang and Mingming Jing

**Abstract** The problem of computing a migration plan among the store devices for moving the data from current configuration to the target configuration is called data migration problem. There are some reasons to cause the data migration, the reordering of the data combination, the system's load balancing, and the change of use mode. In this paper, we ignore the network speed and transfer speed. And only consider the situation that each store device can be used as sender or receiver in the transfer process at the same time. We develop two algorithms based on the multi-graph edge coloring problem: complete decomposition algorithm and the Greedy of maximum degree and weigh match algorithm. They can be used in different circumstances and has better performance, they have better parallelism in the data migration, and consume less time in the process of data migration.

**Keywords** Data migration · Edge coloring · Max-degree and time greedy match

### 86.1 Introduction

For the modern large-scale storage systems (such as large-capacity disk) device, it consists of several disks connected using a dedicated network, there are two key reasons to influence the system performance: load balancing and the transfer

---

G. Huang (✉) · M. Yang

The School of Electronic and Informatin Engineering, Xi'an Jiaotong University,  
Xian, Shannxi, China

e-mail: huanggangfeng@stu.xjtu.edu.cn

M. Yang

e-mail: msy@mail.xjtu.edu.cn

M. Jing

Department of Teaching and Research, PLA Border Defense Academy, Xian,  
Shannxi, China

e-mail: jyy20090123@163.com



speed. Unfortunately, the layout of the data in store devices is not static, in some special circumstances, such as the system upgrade, device addition and removes and so on. The layout of data will change. So it is inevitable to design a data migration plan, whether it is a cyclical change or a new change, the data must be migrated from the old configuration to the new configuration. To minimize the impact of data migration, this migration plan must be as efficient as possible, especially for large data objects (Gigabytes units), the data migration need a few days.

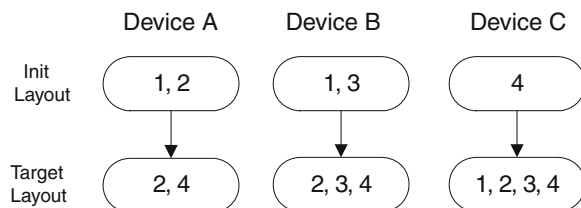
In this paper, we only consider the half-duplex communication model and ignore the new data layout change in the period of the data migration. We develop two algorithms using external disks, and our data migration is running as a background process, can perform an offline migration or less when the user requests (night and weekend).

## 86.2 Models and Assumptions

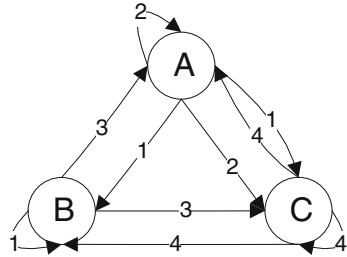
The formal description of the *data migration problem* is as follows: data item  $i$  resides in a specified (source) subset  $S_i$  of disks, and needs to be moved to a (destination) subset  $D_i$ . In other words, each data item that initially belongs to a subset of disks needs to be moved to another subset of disks [1]. (We might need to create new copies of this data item and store it on an additional set of disks.) See Fig. 86.1 for an example. If each disk had exactly one data item, and needs to copy this data item to every other disk, then it is exactly the problem of gossiping.

For Fig. 86.1, we can get every item's source device and destination device, for example, item 1 is located in disk A and B in the initial layout and in the target layout C; we need to create a copy of the item in disk C. Using this method, we can get all item's source device set and destination device set. With these source sets and destination sets, we are led to describe the input to our problem as a directed multi-graph  $G = (V, E)$  without self-loops that we call the demand graph. Each node in the demand graph corresponds to a storage device, and each directed edge  $(u, v) \in E$  represents an object that must be moved from storage device  $u$  (in the initial configuration) to storage device  $v$  (in the final configuration). An example of how a demand graph is defined based on an initial and goal configuration is given in Fig. 86.2.

**Fig. 86.1** The init layout and target layout of devices



**Fig. 86.2** Data migration graph



Most variants one can imagine on this problem are NP-complete [2]. The migration problem for networks of arbitrary topology is NP-complete even if all objects are the same size and each device has only one object that must be moved off of it. The problem is also NP-complete when there are just two storage devices connected by a link, if the objects are of arbitrary sizes.

The time data migration need depends on many factors: network speed, the parallelism of the migration process, the data size of the object. For most storage devices, it only can work as one transmission side at the same time, but some devices are full duplex. In order to make the research simple, before discuss the data migration algorithms, we will make some assumptions:

1. Every device has enough free space for the data migration.
2. Each device can be used as only one object (sender or receiver, but not both) at the same time.
3. The store system has complete network.

With these three assumptions, we will find that the data migration problem in this form is as same as the problem multi-graph edge coloring problem.<sup>1</sup> We can use the model of graph edge coloring as our research model [3]. Our goal is to use minimal color to cover all the edges of the graph.

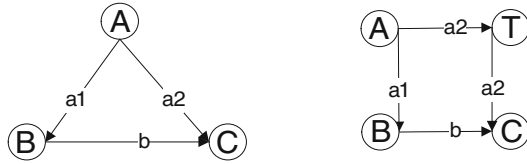
## 86.3 Research Method

### 86.3.1 Use Bypass Node to Improve the Parallelism

First, we explore a variant of the coloring algorithm, by using the bypass node, data movement to reduce the number of colors required. In the indirect data migration plan, a data object may first be sent to a temporary intermediate storage device and it is finally sent to the destination node. This can increase the parallelism of the migration process [4].

<sup>1</sup> **Definition** The chromatic index of a graph is the number of colors required to color the edges of the graph in such a way that no two adjacent edges have the same color.

**Fig. 86.3** Use bypass node to reduce the number of colors. On the *left* side, we need three colors, and on the *right* side, by using the device T as the bypass node, we only need two colors



**Proposition 1** Use bypass node to improve the parallelism (Fig. 86.3).

Before introduce out algorithm, there are two theorems for out algorithms.

**Theorem 1** A connected undirected graph has an Euler cycle  $\Leftrightarrow$  each vertex is of even degree [5].

**Theorem 2** Hall's marriage theorem

A class contains n girls and n boys. Each girl likes k boys and does not like others. Each girl can be paired up with a boy that she likes [6].

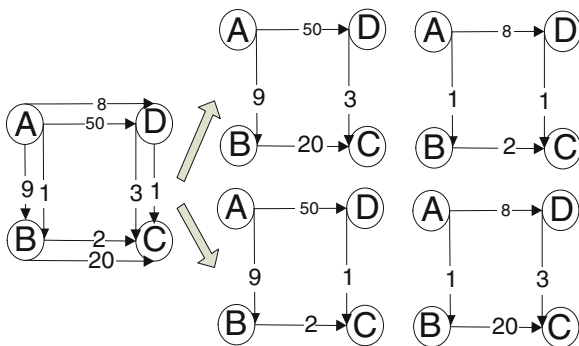
**Algorithm 1** Complete Decomposition Algorithm

1. In graph  $G(V, E)$ , if the maximum degree is  $\Delta'$ , add edge, make sure every vertex's degree is  $2\Delta'/2$ , get a regular graph  $G'$ ;
2. Find an Euler-tour of  $G'$ , And to reassign the direction in accordance with the direction of the Euler-tour [7];
3. Set up a bipartite graph with the Euler-tour of  $G'$ , with a representative of each node in the graph on both sides. Make sure all edges going from left to right;
4. Find a match of the bipartite graph, remove the edges from  $G'$ , and use it build a new graph  $G''$ , this  $G''$  will include all the vertexes, and all the vertexes' degree is 2;
5. In the  $G''$ , if exists exist cycle with odd length path, add an bypass node w, and find out an edge (u, v) in the cycle, and then use edge (u, w) and (w, v) replace the edge (u, v);
6. For every components of  $G''$ , use A/B alternately label the edge, for the sub-components with bypass node, make sure the edge send to the bypass node is marked as A [8];
7. At first, execute the edge with flag A, and then execute the edge with label B;
8. If there are still edges in graph  $G'$ , go back to step 2;

### 86.3.2 Use Greedy Match Algorithm to Cut Down the Time Needed

Sometimes, the time every edge needed is different. In the process of migration, for parallel edge, the consume time is the maximum of value.

**Fig. 86.4** Left side is the original graph, the edge value is the time it needed, we can use different methods decomposition it, but the total time it consumed is different



To minimize the total time about move data on the disk, it is a good idea to make sure the objects with high weight execute at the same time [9] (Fig. 86.4).

Yet there is a problem: how to make the balance about bypass node number and the time spent by the migration? We develop another algorithm for this situation

**Algorithm 2** Greedy Match Algorithm

1. In graph  $G(V, E)$ ,  $\Delta'$  is the maximum degree, for each vertex with degree  $\Delta'$ , find the edge with biggest weight, delete this edge from  $G$  and added it into  $G'$ , if the edge already exists in  $G'$ , then skip it. (after this step,  $G'$  will include all the vertex with degree  $\Delta'$ , and each vertex has a edge connected to it);
2. Find the edge  $(u, v)$  with maximum weight in  $G$ , and if in graph  $G'$ , the degree of the  $u, v'$  is less than 2, add  $(u, v)$  into  $G'$  in and deleted it from the  $G$ . Cycle doing this step until no edge math the condition [10];
3. In Graph  $G'$ , check whether exist cycle with odd length path, if exists, add an bypass node  $w$ , and find out an edge  $(u, v)$  with minimum weight, and then use edge  $(u, w)$  and  $(w, v)$  replace the edge  $(u, v)$ ;
4. For every components of  $G'$ , use A/B alternately label the edge, for the sub-components with bypass node, make sure the edge send to the bypass node is marked as A;
5. At first, execute the edge with flag A, and then execute the edge with label B;
6. If there are still edges in graph  $G$ , go back to step 1;

**86.4 Results and Analysis**

In order to validate our algorithm, we need to calculate the total time they need for the data migration, and verify that the required number of bypass node decomposition algorithm and the greedy algorithm.

We tested all of these algorithms on two types of multi-graphs:

1. Load-balance graph, the time every edge need is very similar.
2. Non load-balance graph, for this type graph, the time every edge need is quite different.

### 86.4.1 Load-Balance Graph

A data migration graph can be generated from any two configurations of a set of objects on a set of nodes in a network. To generate the *Load-Balancing* graphs [11], we used a method of generating sequences of configurations of objects which might occur in a real world system.

We will get 100 graphs, for each degree  $D(i)$ , we generate 10 different graphs with different nodes number  $N(i)$ ,

$$N(i) = 100 + 200 * (i - 1) \quad i = 1, 2 \dots 10$$

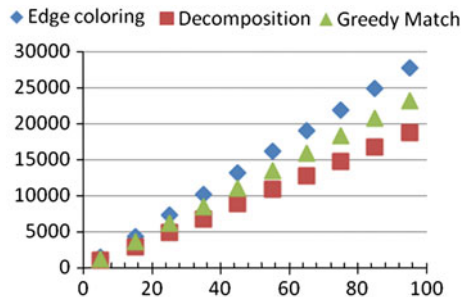
$$D(i) = 5 + 10(i - 1), \quad i = 1, 2 \dots 10$$

The number of nodes in each graph is from 100 to 1 for the graphs in all sets except the third in which most graphs has around 300 nodes. The edge degree for each graph varies from about five for most graphs in the third set to around 65 for most graphs in the fourth set (Fig. 86.5).

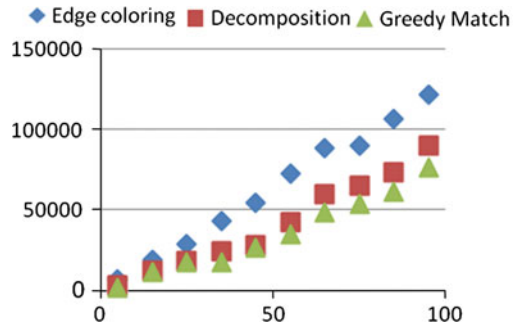
### 86.4.2 Non-Load-Balance Graph

For the situation that all edge with the extreme disparity time value, decomposition has best performance (Fig. 86.6).

**Fig. 86.5** All edge with the similar time value, decomposition has best performance



**Fig. 86.6** The total time of algorithms



## 86.5 Conclusions

In this paper, we studied the data migration problem. Our objective is to find a migration plan among the storage devices. Our experiments indicate that the weighted matching heuristic with some randomness does very well. It converts the data migration problem to the edges of the graph coloring problem [12] and use bypass nodes to improve the parallelism of data move process. We develop two algorithms for different circumstances. We have empirically tested the algorithms and have found that there are some quite good performances in practice.

## References

1. Khuller S, Kim Y, Wan Y-C (2003) Algorithms for data migration with cloning. In: 22nd ACM symposium on principles of database systems (PODS), pp 27–36
2. Anderson E, Hall J, Hartline J, Hobbes M, Karlin A, Saia J, Swaminathan R, Wilkes J An experimental study of data migration algorithms. In: Workshop on algorithm engineering
3. Zhang C-Q (1997) Integer flows and cycle covers of graphs. Marcel Dekker Inc, New York
4. Bunke H, Jiang XY, Abegglen K et al (2002) On the weighted mean of a pair of strings. *Pattern Anal Appl* 5(5):23–30
5. Golubchik L, Khanna S, Khuller S, Thurimella R, Zhu A (2000) Approximation algorithms for data placement on parallel disks. In: Proceedings of ACM-SIAM SODA, 2000
6. Miao L, Liu G (2002) Edge covered coloring and fractional edge covered coloring. *J Syst Sci Complexing* 15(2):187–193
7. Nakano S-I, Nishizeki T (1994) Edge-coloring problem for graphs. *Interdiscip Inf Sci* 1(1):19–32
8. Wu B, Lawless D, Bisbal J, Grimson J, Wade Vincent, Richardson R, O’Sullivan D The mechanisms of DAA and chrysaliser for butterfly methodology. Technical report, Department of Computer Science, Trinity College, Dublin
9. Liben-Nowell D (2001) Gossip is synteny: incomplete gossip and an exact algorithm for syntenic distance. In: Proceedings of ACM-SIAM SODA, pp 177–185
10. Narayanan D, Thereska E, Donnelly A, Elnikety S, Rowstron A (2009) Migrating server storage to SSDs: analysis of tradeoffs. In: The 4th ACM European conference on computer systems (Eurosys), New York

11. Verma A, Koller R, Useche L, Rangaswami R (2010) SRC map: energy proportional storage using dynamic consolidation. In: The 8th USENIX conference on File and storage technologies (FAST), San Jose, Feb 2010
12. Kang S, Park H, Yoo C (2011) Performance enhancement of I/O scheduler for solid state devices. In: IEEE international conference on consumer electronics (ICCE), Jan 2011

# Chapter 87

## Medical Image Fusion Algorithm Based on the Laplace-PCA

Pengtao Zhao, Gang Liu, Cen Hu, Huang Huang and Bing He

**Abstract** Medical image fusion processing, as an indispensable part of the modern medical treatment, has been used widely in clinic medicine. The Paper firstly describes the Gauss decomposition of the medical image and the establishment of the Laplace pyramid decomposition images, then fusing medical image using PCA fusion criteria, finally obtaining the fusion image based on the Laplace pyramid image reconstruction. The experimental results show that the algorithm can be complementary information of the CT image and MR image highlights, and has a good fusion effect.

**Keywords** Gauss pyramid · Laplacian pyramid · PCA · Image fusion

### 87.1 Introduction

Image fusion is that integrated and extracted two or more of the original image information, to obtain a more accurate, comprehensive and reliable image of the same scene or target, to make it more suitable for the human eye perceives or computer subsequent treatment [1, 2]. Image fusion is an important branch of the multi-sensor data fusion, and it takes full advantage of multiple fusion of redundant information contained in the image and complementary information, is different from the general sense of the image enhancement [3]. Medical image fusion processing, as an indispensable part of the modern medical treatment, has been used widely in clinic medicine, such as plan design, program implementation and efficacy assessment of disease diagnosis, surgery and radiation therapy [4].

---

P. Zhao (✉) · G. Liu · C. Hu · H. Huang · B. He  
College of Science, The Second Artillery Engineering University, Xi'an 710025, China  
e-mail: zx103786260@163.com



Pyramid image fusion processing algorithm was first proposed by Burt and Adelson [5], and is a multi-scale, multi-resolution method. Image fusion algorithm based on pyramid decomposition is conducted in different scales and spatial resolution and different decomposed layer, respectively [6]. Compared with the simple image fusion algorithm, it can obtain better fusion effect, and used in broader occasions [7]. The algorithm of image fusion criterion has pixel gray value weighted average, maximal pixel selection and minimal pixel selection [8, 9]. The key of pixel gray value weighted average method lies in the determination of the weighting coefficients, and PCA (Principal Component Analysis) is used to determine the weighting coefficients.

The basic idea of this paper is that retaining the image boundary information using Laplace pyramid decomposition, determining the weighting coefficients using PCA algorithm, so as to achieve more objective medical image fusion. The Paper firstly describes the Gauss decomposition of the medical image and the establishment of the Laplace pyramid decomposition images, then fusing medical image using PCA fusion criteria, finally obtaining the fusion image based on the Laplace pyramid image reconstruction.

## 87.2 Laplace Pyramid Decomposition Principle

Laplace pyramid image processing is obtained in the the Gauss pyramid image basis, and therefore has medical image Gauss pyramid decomposition firstly.

### 87.2.1 Gauss Pyramid Decomposition

Source image  $G_0$  is the Gauss bottom of the pyramid (layer 0). The first layer of the Gauss pyramid is obtained through the Gauss low-pass filtering and interlaced sampling separation column of the original image, then obtained the second layer of the Gaussian pyramid by low-pass filtering and lowersampling the first layer image. Gaussian pyramid image can be obtained by repeating the above process. Gauss pyramid image  $G_l$  (layer  $l$ ):

$$G_l = \sum_{m=-2}^2 \sum_{n=-2}^2 \omega(m, n) G_{l-1}(2i + m, 2j + n) \quad (87.1)$$

$$0 < l \leq N, 0 < i \leq C_l, 0 < j \leq R_l$$

wherein  $N$  represents the number of layers of the Gauss pyramid top;  $C_l$  represents the number of columns of the Gauss pyramid layer  $l$ ;  $R_l$  represents the number of rows of the Gauss pyramid layer  $l$ ;  $G_{l-1}$  is the Gauss pyramid layer  $l-1$ ;  $\omega(m, n)$  is

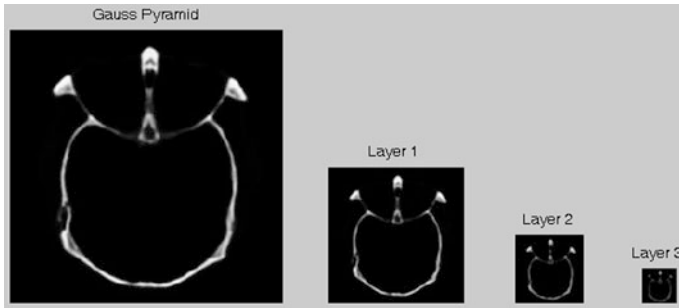


Fig. 87.1 Gauss pyramid image

the  $5 \times 5$  window function of Gauss low pass characteristics, its expression is as follows:

$$\omega(m, n) = \frac{1}{256} \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} \tag{87.2}$$

$G_0, G_1, \dots, G_N$  constitute Gauss pyramid, where  $G_0$  is the bottom of the pyramid (it is the original image),  $G_N$  is the top level of the pyramid. This layer image is the result of previous layer images Gaussian low-pass filtering and making of interlaced and sampling every column down 2. The size of this layer image is 1/4 of the previous layer image. Gauss pyramid decomposition of the image is obtained as shown in Fig. 87.1.

### 87.2.2 Laplace Pyramid Decomposition

$G_l$  is interpolated zoom, and obtained enlarged image  $G_l^*$ . The size of  $G_l^*$  is the same as the size of  $G_{l-1}$ . The expression of  $G_l^*$  is as follows:

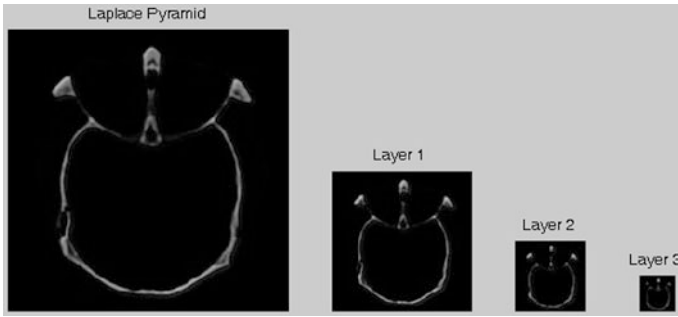
$$G_l^*(i, j) = 4 \sum_{m=-2}^2 \sum_{n=-2}^2 \omega(m, n) G_l \left( \frac{i+m}{2}, \frac{j+n}{2} \right) \tag{87.3}$$

$0 < l \leq N, 0 < i \leq C_l, 0 < j \leq R_l$

wherein  $G_l \left( \frac{i+m}{2}, \frac{j+n}{2} \right) = \begin{cases} G_l \left( \frac{i+m}{2}, \frac{j+n}{2} \right) & \text{When the } \frac{i+m}{2}, \frac{j+n}{2} \text{ is an integer} \\ 0 & \text{Others} \end{cases}$ .

The expression of Laplace pyramid image is as follows:

$$\begin{cases} LP_l = G_l - G_{l+1}^* \\ LP_N = G_N \end{cases} \tag{87.4}$$



**Fig. 87.2** Laplace pyramid image

wherein,  $N$  is the number of layers of the Laplace pyramid top;  $LP_l$  is the Laplace pyramid decomposition  $l$ -layer image.

$LP_0, LP_1, \dots, LP_N$  constitute the Laplace pyramids. Each layer image is the difference between a Gauss pyramids present layer and its high interpolated image. This process is equivalent to a band-pass filter, so Laplace pyramid is also known as the bandpass pyramid decomposition. Laplace pyramid decomposition of the image shown in Fig. 87.2. The Laplace pyramid image can express the the boundary characteristic information of the original image.

### 87.2.3 Reconstruct the Original Image by the Laplace Pyramid

Reconstructed image is done by the Laplace pyramid image according to the formula (87.4). Its expression is as follows:

$$\begin{cases} G_N = LP_N \\ G_l = LP_l + G_{l+1}^* \end{cases} \quad (87.5)$$

According to formula (87.5), layer by layer from top to bottom starting from the top of the Laplace pyramid, its corresponding Gauss pyramid can be recursived, and, ultimately, the original image can be obtained.

### 87.3 Image Fusion Based on Laplace-PCA

#### 87.3.1 Laplace Image Fusion Principle

The purpose of image Laplace pyramid decomposition is the decomposition of the source image to a different spatial frequency band. The fusion process is carried out separately in each of the spatial frequency layer, so that the features and details may be on different frequency bands for different decomposed layer, aiming to use different fusion operator to highlight specific band features and details, that may be fused together from different image features and details is shown in Fig. 87.3, in which the image M, N are two source images to be fused, F is the fusion image. Fusion algorithm of multiple images can be and so on.

#### 87.3.2 PCA Fusion Criterion

In practical problems, it involves a large number of variables, and in many cases, There is some correlation between variables and these variables are provided overlap information to a certain extent. People want to be the “transformation” of these variables. Very few new complementary variables reflects most of the information provided in the original variable. The new variable achieves the purpose of problem-solving, which is the idea of principal component analysis (PCA). Image PCA algorithm is that after discarding a minor constituent of the poor correlation, determining the weight of being fused image based on the main component. Then weighted fusing based on its weight. Fused image is the best approximation in the statistical sense of the original image.

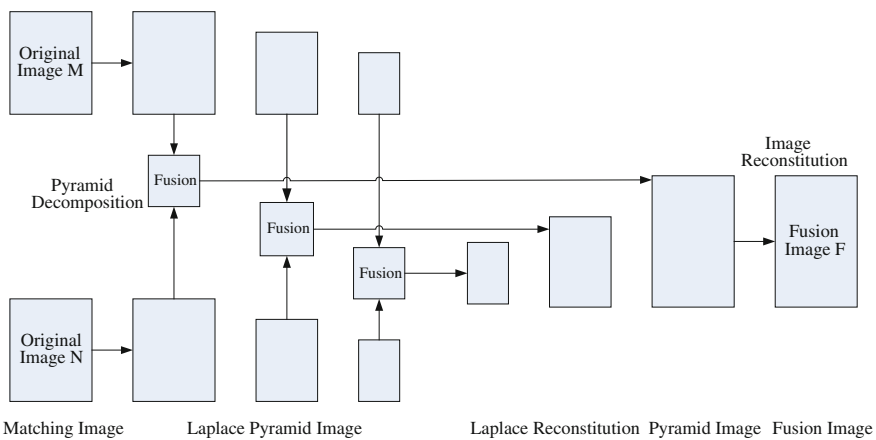


Fig. 87.3 Block diagram of image fusion algorithm based on Laplace pyramid decomposition

PCA algorithm process is as follows:

1. The column vector is row or column priority components of the image, and calculating covariance;
2. Striking a feature vector based on the covariance matrix, and determining the feature vectors corresponding to the first principal component;
3. In accordance with the feature vectors corresponding to the first principal component, weights is determined.

Suppose there are two images A, B. The feature vector of its first principal component is  $(x, y)^T$ , the weight of image A is  $\frac{x}{x+y}$ , the weight of image B is  $\frac{y}{x+y}$ .

### 87.3.3 Laplace-PCA Fusion Step

Flow chart of medical image fusion algorithm based on Laplace-PCA is shown in Fig. 87.4.

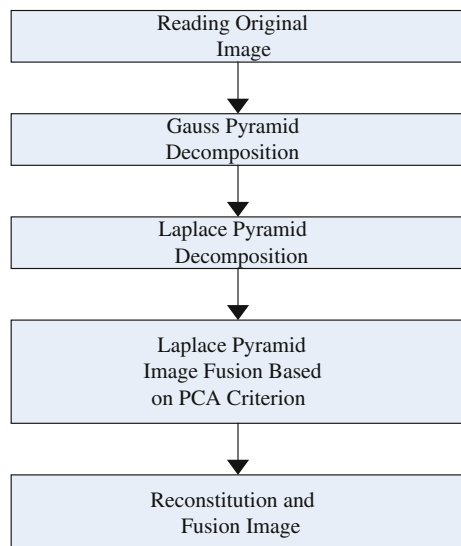
The steps are as follows:

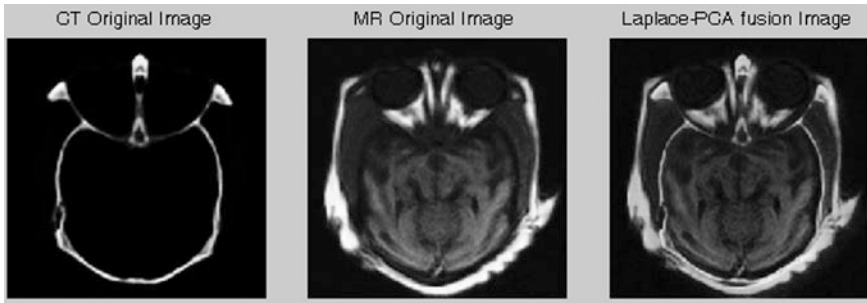
Step 1: Each piece of source image are done to the Gauss pyramid decomposition;

Step 2: Each of Laplace pyramid were constructed for each piece of source images;

Step 3: Weight is determined based on the the PCA fusion criterion, then to be fused Laplace pyramid;

**Fig. 87.4** Flow chart of image fusion algorithm based on Laplace-PCA





**Fig. 87.5** Fusion results based on Laplace-PCA

Step 4: After fusion Laplace pyramid image reconstructing, obtaining the final fusion image.

## 87.4 Conclusion

Fusion algorithm MATLAB simulation results based on Laplace-PCA shown in Fig. 87.5. Black in CT images said the low-absorption region, namely, low-density areas, namely ventricle; while white indicates high absorption region, namely high density areas, namely the skull. Intracranial soft tissue can be well manifested in MR images. Two types of image is with information complementary. Image fusion algorithm based on Laplace-PCA is able to highlight the respective complementary information fusion.

## References

1. Han C, Zhu H, Duan Z (2010) Multi-source information fusion, 2nd edn. Tsinghua University Press, Beijing (in Chinese)
2. Marcin D, Michal C, Pawel M (2009) A new image fusion method for estimating 3D surface depth. Springer, Berlin, pp 90–99
3. Ishita D, Bhabatosh C, Buddhajyoti C (2006) Enhancing effective depth-of-field by image fusion using mathematical morphology. *Image Vis Comput* 24:1278–1287
4. Peng S (2006) Medical image registration and fusion method based on multi-scale analysis. The Degree of Doctor of Philosophy in Engineering, Huazhong University of Science & Technology, Wuhan (in Chinese)
5. Burt PJ, Adelson EH (1983) The Laplacian pyramid as a compact image code. *IEEE Trans Commun* 31(4):532–540
6. Chen H, Wang Y (2009) Research on image fusion algorithm based on Laplacian pyramid transform. *Laser Infrared* 39(4):339–442 (in Chinese)

7. Jan A, Hiệp Q, Bart G, Aleksandra P, Wilfried P (2011) Augmented Lagrangian based reconstruction of non-uniformly sub-Nyquist sampled MRI data. *Sig Proc* 91:2731–2742
8. Zhao P, Bai Z, Fan W (2010) Research of fast image matching based on PCA. *Appl Electron Tech* 4:132–135
9. Pan Y, Sun Q, Xia D (2011) Image fusion framework based on PCA decomposition. *Comput Eng* 37(13):210–212

# Chapter 88

## Filter Parameter Estimation in Non-Local Means Algorithm

Hong-jun Li, Wei Hu, Zheng-guang Xie and Yan Yan

**Abstract** In this paper, improvements to the Non-local Means (NL-Means) algorithm introduced by Buades et al. are presented. The filtering parameter is unclearly defined in the original NL-Means algorithm. In order to solve this problem, we calculated filtering parameter by the relation of noise variance, and then proposed a noise variance estimate method. In this paper, noisy image is transformed by wavelet. The wavelet coefficients in each sub-band can be well modeled by a Generalized Gaussian Distribution (GGD) whose parameters can be used to estimate noise variance. The simulation results show that the noise variance estimate method is not only exact but also makes the algorithm adaptive. The adaptive NL-Means algorithm can obtain approximately optimal value, and need less computing time.

**Keywords** Non-local means algorithm · Filtering Parameter · Generalized Gaussian Distribution · Wavelet Domain

### 88.1 Introduction

In recent years, patch-based methods have drawn a lot of attention in the image processing community. Buades et al. have proposed the Non-local Means denoising algorithm [1]. That paper proposed a very elegant nonlocal denoising method

---

H. Li (✉) · Z. Xie

School of Electronic Information Engineering, Nantong University, 226019 Nantong, China  
e-mail: lihongjun103@126.com

W. Hu

Nantong University, 226019 Jiangsu, Nantong, China

Y. Yan

School of Computer Sciences and Technology, Nantong University, 226019 Nantong, Jiangsu, China



shown to propose a state-of-the-art result. The NL-Means restores the original image by considering Non-local neighborhoods of a given pixel.

The aim of this paper is to discuss the choice of parameters of the standard NL-means filter. To discuss the tuning of parameters of the NL-means algorithm, we interpret this choice as a bias-variance dilemma. Contrary to the approach of [2], which also relies on a bias variance analysis, we mainly focus on the choice of the smoothing parameter rather than on the search window. This choice should be made depending on the regularity of the image, a notion that is to be defined. The study of [3] discussed the influence of two important parameters on this algorithm: the size of the searching window and the weight given to the central patch.

The Filter parameter in [1] is searched in oracle, and set an optimal value as the filtering parameter in NL-Means algorithm. The author mentioned that the value of Filter parameter is correlated to image noise variance, but the definition of Filter parameter is unclearly. In this paper, we first present an overview of the NL-Means algorithm introduced in [1]; also analyze the relation between Filter parameter and noise variance. Then we estimate the noise variance by the parameters of GGD model. Finally, we use the estimated Filter parameters to solved traditional NL-Means algorithm drawbacks.

## 88.2 NL-Means Image Denoising Method

In this section, a brief overview of NL-Means algorithm is introduced. Given a discrete noisy image  $v = \{v(i)|i \in I\}$ , the estimated value  $NL[v](i)$  is computed as a weighted average of all the pixels in the image,

$$NL[v](i) = \sum_{j \in I} w(i,j)v(j) \quad (88.1)$$

where the family of weight  $\{w(i,j)\}_j$  depend on the similarity between the pixels  $i$  and  $j$ , and satisfy the usual conditions  $0 \leq w(i,j) \leq 1$  and  $\sum_j w(i,j) = 1$ .

The similarity between two pixels  $i$  and  $j$  depends on the similarity of intensity gray level vectors  $v(N_i)$  and  $v(N_j)$ , where  $N_k$  denotes a square neighborhood of fixed size and centered at a pixel  $k$ . This similarity is measured as a decreasing function of the weighted Euclidean distance,  $\|v(N_i) - v(N_j)\|_{2,a}^2$ , where  $a$  is the standard deviation of the Gaussian kernel. The application of Euclidean distance to noisy neighborhoods raises the following equality.

$$E\|v(N_i) - v(N_j)\|_{2,a}^2 = \|u(N_i) - u(N_j)\|_{2,a}^2 + 2\sigma^2 \quad (88.2)$$

This equality shows the robustness of the algorithm since in expectation the Euclidean distance conserves the order of similarity between pixels. So the most

similar pixels to  $i$  in  $v$  also are expected to the most similar pixels  $i$  in  $u$ . The weights express the amount of similarity between the neighborhoods of each pair of pixels involved in the computation.

### 88.3 Obtain the Value of Filter Parameter

In [1] Buades obtained the filtering parameter  $h$  by experiments, and found that the value of the filtering parameter  $h$  is between  $10 \times \sigma$  and  $15 \times \sigma$ . Buades made an error, he wanted to say  $h^2 = 10 \times \sigma$  but he said  $h = 10 \times \sigma$ . In [4] the author discussed the relation between filtering parameter  $h$  and noise variance, and gave a relation between them. However, too many parameters are considered in estimation and new parameters lead the relation uncertainly. The [4] proposed that  $h$  parameter and noise variance are linear correlative.

In Eq. (88.2), the weight function is Gaussian kernel with standard deviation  $h/\sqrt{2}$  acts as a filtering parameter. On the other hand, under the circumstances that the noise presents as a Gaussian distribution, the pixel and neighborhood pixels have the same value of gray. But the added Gaussian white noise will make them have different value of gray. Gaussian white noise will cause the difference between current pixels which can be calculated by Euclidean difference. So the value of filtering parameter is approximate  $\sqrt{2}$  times of noise sigma. It is difficult to obtain the value of filtering parameter accurately, so we chose the approximate one to substitute. The filtering parameter can be calculated by noise variance estimation. We show how to estimate in next section.

### 88.4 Noises Estimate

Many image denoising algorithms assume that the noise level is known prior; and the algorithms are adapted to the amount of noise instead of using fixed parameters. In 1993, Olsen [5] gave a complete description and comparison of six earlier estimation algorithms. They are classified into two different approaches: filter-based and block-based. In filter-based methods, the noisy image is firstly filtered by a low-pass filter to suppress the image structures. Then the noise variance is computed from the difference between the noisy image and the filtered image. The main difficulty of filter-based methods is that the difference image is assumed to be the noise but this assumption is not true as images have structures or details. In block-based methods, images are tessellated into a number of blocks. The noise variance is then computed from a set of homogeneous blocks. The main issue of block-based methods is how to identify the homogeneous blocks.

In this section we used the image block method in the wavelet domain, and proposed a novel image noise estimate method. In the wavelet domain, the most simple method about noise variance estimation is calculated by diagonal sub-bands

coefficients. The noise variance estimate method proposed by Donoho and Johnstone [6] is expressed as follows:  $\sigma_n = MAD/0.6745$ ; Where MAD is middle value of diagonal sub-bands coefficients. In methods of [6], it is inappropriate to take the entire coefficients in HH sub-band to estimate noise variance without considering the coefficients also contain the edge information in direction 45 and 135°. So these methods are no suitable to estimate noise variance.

The threshold proposed by Chang [7] is derived in a Bayesian framework, and the prior used on the wavelet coefficients is the GGD widely used in image processing applications. The proposed Bayesian risk minimization is sub-banded dependent. Given the signal being generalized Gaussian distributed and the noise being Gaussian, via numerical calculation a nearly optimal threshold defined as:  $T_B(\sigma_s) = \sigma^2/\sigma_s$ ; where  $\sigma^2$  is the noise variance, and  $\sigma_s^2$  is the signal variance. This threshold only on the standard deviation and not on the shape parameter, it may not yield a good approximation for values of shape parameter, and the threshold may need to be modified to incorporate shape parameter. Our method is based on the Bayes-shrink threshold and takes the two parameters of GGD model into account.

#### 88.4.1 Wavelet Coefficients Modeled by GGD Model

Mallat [8] proved that the histogram of wavelet coefficients can be well fitted by GGD model. Define GGD,  $p(x; \alpha, \beta)$ ,

$$p(x; \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp(-|x|/\alpha)^\beta \quad (88.3)$$

$$\alpha = \sigma_x \sqrt{\frac{\Gamma(1/\beta)}{\Gamma(3/\beta)}} \quad \beta > 0 \quad (88.4)$$

where,  $\sigma_x$  is the standard deviation of the signal and  $\Gamma(\cdot)$  is the Gamma function;  $\alpha$  is the scale parameter and  $\beta$  is the shape parameter. The value of  $\beta$  determines the decay rate of the probability density function. Then our purpose is to obtain the scale and shape parameters.

#### 88.4.2 Analyze the Parameter of GGD Model

Zhang [9] find that many well-known types of image distortions lead to significant changes in wavelet coefficient histograms, and measurement is based on the parameters of generalized Gaussian model. The paper introduce a new way to calculate the distortion of image, so there are some relation on the image distortion and the parameters of generalized Gaussian model indeed. So to find out a proper relationship of them is most important. In the previous work, Do and Vetterli [10]

used GGD model to fit distribution of wavelet coefficients, and proved shape parameter is among [0.7 2.0]. When noise variance increases, shape parameter changes smoothly. When shape parameter is among [0.7 2.0], the relation of scale parameter and the standard deviation of the signal is approximately linear. On the experiment, when noise variance increases, shape parameter changes smoothly. So when add noise increase, and research the point, shape parameter can not well calculate by the standard deviation of the signal. So we need to analysis the relationship of shape parameter and noise variance.

We used several images for test under different noise variance, and obtained the sequence value of  $\alpha$  on different types of images. In order to obvious the relationship of scale parameter and noise variance deeply; we used experiments to find out the relationship of them. It is an interesting observation that the reciprocals of scale parameter changes smoothly when noise variance research the point, it is similarly the character of shape parameter. The value of  $1/\alpha$  is reduced dramatically when the value of noise variance is small, and then keeps a fixed value when noise variance reach a certain value. As the noise variance increase the value of  $1/\alpha$  changes small. When  $\sigma = 20$ , it keeps among [0.035 0.040]. When the noise sigma is bigger than 20, the value of  $\beta$  is bigger than 1 in most images. We find that in the situation of low density noise; if the value of  $1/\alpha$  is smaller than a certain value, the image has the character of abundant information and detail structure.

### 88.4.3 Noise Estimate Method

We found that the scale parameter and shape parameter are closely related to the standard deviation of noise, and have some superior characters. Our estimation method is consider both the scale parameter and shape parameter. We found the shape parameter increases a little, when standard deviation of noise increases. We observed that a GGD with the shape parameter ranging from 0.5 to 1 can adequately describe the wavelet coefficients of a large set of natural images. The detail of estimate method is show in the follow steps. 1. Wavelet transformation in noisy image; 2. Used GGD to model wavelet coefficients probability histogram in each scales and directions; 3. Donoho noise variance estimate method; 4. Image noise estimate, using the scale and shape parameter: (a) When  $\beta \geq 1$ ,  $\hat{\sigma} = \sigma_n \times (\sigma_n/\sigma_x)^{(2-\beta)/\beta}$ ; (b) when  $\beta \leq 1$  and  $\alpha^{-1} \leq 0.2$ ,  $\hat{\sigma} = \sigma_n \times (\sigma_n/\sigma_x)^{(1-\beta)/\beta}$ ; (c) when  $\beta \leq 1$  and  $\alpha^{-1} > 0.2$ ,  $\hat{\sigma} = \sigma_n \times (\sigma_n/\sigma_x)^{(1+\beta)/\beta}$ .  $\sigma_x$  is considering all coefficients in image, so that the value is bigger than  $\sigma_n$  in most condition. The value of  $\sigma_n/\sigma_x$  is a little smaller than 1. In high density noise,  $\sigma_n$  is approximate to the noise value. So we set  $\sigma_n$  as a reference value. In low density noise, we use scale parameter to sort different kind of images mentioned above. If scale parameter smaller than 0.2, the image has complex structure and abundant details. In this situation,  $\sigma_n$  is bigger than the added standard deviation of noise.

## 88.5 Adaptive Non-local Means Algorithm

### 88.5.1 Filter Parameter $h$ Estimate

We can calculate the filter parameter  $h$  by the noise variance method proposed above. In Table 88.1, we compare key parameters in original NL-Means algorithm and the method we proposed. Our method can estimate noise variance in good condition, and compute the  $h$  filtering parameter in low computational complexity. The complexity is 10 % of the original method. The proposed value of  $h$  is approximate to the optimal value, and the Peak Signal to Noise Ratio (PSNR) is acceptable.

### 88.5.2 Adaptive Non-local Means Algorithm

We used parameter of GGD to obtain the noise variance, and then used the noise variance to estimate parameter  $h$ . It makes the method adaptive to preprocessing image instead of searching an optimal filtering parameter  $h$ . When image is decomposed by the wavelet, the size of image is reduced half. If the decomposition scale increases, the size will be reduced. Also the computational complexity is reduced quickly. So our algorithm not only solved the set of filtering parameter  $h$ , but also reduced the computation complexity. Figure 88.1 gives the denoising results about different algorithms; Visual quality of our method outperforms the Non-local means method [1], and better the method [11] with different value of filter parameter.

**Table 88.1** Compare with different method

$\sigma$	10	20	30	40	50
<i>Method in [1] using oracle</i>					
Sigma([1])	10	20.4	30.9	41.2	51.0
Time(S)	480	462	473	479	461
h(Optimal)	14.1	28.6	43.2	57.7	71.4
PSNR(Optimal)	31.5	28.7	26.8	25.2	24
<i>Method in [1] using <math>h^2 = 10 \times \sigma</math></i>					
h([1])	10	14.3	17.6	20.3	22.6
Time(S)	47	48	47	45	44
PSNR	31.4	27.8	24.0	20.8	18.3
<i>Method proposed in this paper</i>					
Sigma(proposed)	10.5	21.1	30.5	41	50.3
h(proposed)	14.7	29.5	42.7	57.4	70.4
Time(S)	42	46	45	48	43
PSNR	31.4	28.5	26.7	25.2	23.9



**Fig. 88.1** Image denoising results based on different algorithms: **a** Original image; **b** noisy image; **c** method in [11] with  $h = 1.2\sigma$ ; **d** method in [11] with  $h = 1.4\sigma$ ; **e** method in [1]  $h^2 = 10\sigma$ , 2007; **f** our method

## 88.6 Conclusions

In this paper, we proposed a noise estimate method to solve the problem about the definition of parameter  $h$  and the estimation of noise variance. We used the parameters of GGD model to estimate the noise variance, also took the noise variance in whole area into account. The classical noise estimate method is the fundamental of our method and support it in theoretic. Noise estimate method proposed in this paper, not only solved the problem about parameter  $h$ , and reduced the computation complexity of traditional Non-local Means algorithm when searching an optimal filtering parameter  $h$ . So the Adaptive NL-Means Algorithm can succeed automatically.

**Acknowledgments** This work was supported by the National Natural Science Foundation of China (NO. 61171077), University Science Research Project of Jiangsu Province (NO. 12KJB510025, NO. 12KJB510026), the Traffic Department of applied basic research project of China (NO. 2011-319-813-510), the science and technology supporting plan (social development) of Jiangsu Province (NO. BE2010686), joint tackle hard-nut problems in science and technology on traffic and transportation industry of state ministry of communications (NO. 2010-353-332-110).

## References

1. Buades A, Coll B, Morel JM (2005) A review of image denoising algorithms, with a new one. *Multiscale Model Simul* 4(2):490–530
2. Kervrann C, Boulanger J (2008) Local adaptivity to variable smoothness for exemplar-based image denoising and representation. *Int J Comput Vis* 79:45–69
3. Salmon J (2010) On two parameters for denoising with non-local means. *IEEE Signal Process Lett* 17:269–272
4. Wang ZM, Zhang L (2009) An adaptive fast non-local image denoising algorithm. *J Image Graph* 14:669–675 (in Chinese)
5. Olsen SI (1993) Estimation of noise in images: an evaluation. *Graph Models Image Process* 55:319–323
6. Donoho DL, Johnstone LM (1994) Ideal spatial adaptation via wavelet shrinkage. *Biometrika* 81:425–455
7. Chang SG, Yu B, Vetterli M (2000) Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans Image Process* 9:1532–1546
8. Mallet S (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans Pattern Anal Machine Intell* 11:674–693
9. Wang Zhou, Simoncelli Eero P (2005) Reduced-reference image quality assessment using a wavelet-domain natural image statistic model, human vision and electronic imaging X. *Proc SPIE* 5666:149–159
10. Do MN, Vetteli M (2002) Wavelet based texture retrieval using generalized Gaussian density and Kullback-leibler distance. *IEEE Trans Image Process* 11:146–158
11. Manjón J, Carbonell J, Lull J, Robles M et al (2008) MRI denoising using non-local means. *Med Image Anal* 12(4):514–523

# Chapter 89

## Image Fusion Using Compressed Sensing in Nonsubsampled Contourlet Transform Domain

Fu Liu

**Abstract** Image fusion algorithm using compressed sensing theory in NonSubsampled Contourlet Transform (NSCT) domain is proposed, NSCT can provide better sparsity than wavelet transform in image transform. After the transform of NSCT, low-frequency coefficients of the image are preserved, only high-frequency coefficients are measured. Fused coefficients are calculated according to different fusion rules in low frequency and high frequency domain. In the reconstruction, OMP algorithm is used to recover the high-frequency coefficients and the image is reconstructed by inverse nonsubsampled contourlet transform. Compared with wavelet compressed sensing algorithms, simulation results demonstrate that the quality of reconstructed image can be greatly improved.

**Keywords** Nonsubsampled contourlet transform • Image fusion • Multi-resolution • Compressed sending

### 89.1 Introduction

Research about multi-sensor information fusion is very popular in recent years, image fusion technology is one of the most important part in this research area. Image fusion can combine information from multiple images of the same scene, and acquire more exact and comprehensive description of the image. So, the technology of multi-sensor image fusion is applied widely on many fields, such as computer vision, remote sensing, intelligent robot, medical image analysis and so on. And the development of multi-sensor image fusion technology is developed very quickly.

---

F. Liu (✉)

School of Information Science and Engineering, Hunan International Economics University,  
Changsha 410205 Hunan, China  
e-mail: femiem@163.com



There are many kinds of fusion algorithms based on transform domain, multi-resolution image fusion, such as Pyramid and multi-resolution wavelet transform [1] are common. Because of the correlation between layer decomposition, image fusion effect of pyramid transform is not very good. Wavelet transform has good properties of time domain and frequency domain and multi-resolution. Image's low frequency information and three high frequency information (horizontal, vertical and diagonal) can be obtained easily. Its fusion effect is better. But wavelet transform suit isotropic objection, for those anisotropic objectives, such as borderline and lineal objection in image, it is not a very good tool.

In order to solve these defects, Contourlet Transform (CT) was put forward by Do and Vetteri [2]. But down-sample and up-sampling in Contourlet Transform made it lack Shift-invariance property. Shift-invariance is very important in image enhance, smooth and fusion. There will be obvious Gibbs phenomenon in image fusion if no property of Shift-invariance. Nonsubsampled Contourlet Transform (NSCT) is made up of a nonsubsampled pyramid and a nonsubsampled directional filter bank, it is similar to contourlet transform, except for most of the excellent property which contourlet has, nonsubsampled contourlet transform also have shift-invariance and higher redundancy.

Conventional fusion methods require some prior knowledge of original images, The recent theoretical results in compressive sensing (CS) [3–5] prove that a sparse or compressible signal can be exactly reconstructed from a small number of nonadaptive linear projections, which is far fewer than the number of samples which the signal is sampled at the Nyquist rate. Thus, CS has the superiority in reducing storage space and computation cost, and simplicity on the hardware side [6–9]. Another key merit offered by CS is that samples can be collected without assuming any priori information about the signal being observed. Owing to these benefits, CS is an attractive scheme for image fusion, and there are many advantages if combining the CS technique with image fusion application [10].

## **89.2 Nonsubsampled Contourlet Transform and CS Theory**

### ***89.2.1 Nonsubsampled Contourlet Transform***

Contourlet transform is composed of Laplacian Pyramid decomposition and directional filter bank (DFB) [11]. The Pyramid decomposes the original image into low frequency sub-band and high frequency sub-band, high frequency sub-band is decomposed into several directions by DFB. Low frequency sub-band is processed like this repeatedly, so multi-resolution and multi-direction decomposition is completed. Because of the excellent properties of anisotropic and directionality, CT allows different direction numbers in different scale, high frequency information can be divided into several directions.

Nonsampled pyramid structure (NPS) and nonsampled directional filter bank (DFB) are used in NSCT to overcome the disadvantage of Gibbs phenomenon caused by lacking Shift-invariance in CT [12, 13], downsample and upsampling are abandoned in NSCT. The NPS is achieved by using two-channel nonsub-sampled 2-D filter banks, Pyramid filter bank is used first to complete multiscale decomposition. The DFB is achieved by switching off the down-samplers/up-samplers in each two-channel filter bank in the DFB tree structure and up-sampling the filters accordingly to complete multi-direction decomposition. Input image make a convolution with a 2 dimension filter mask to obtain the low frequency approximate image, the differentials of original image and low frequency image is high frequency details. Then details are processed by Nonsampled directional filter bank to get the directional details. As a result, NSCT yields better frequency selectivity, regularity, and shift-invariance.

### 89.2.2 Compressive Sensing Theory

According to Shannon sampling theorem, image can be recovered exactly from a set of uniformly spaced samples taken at the Nyquist rate. The Nyquist rate is so high that too many samples are needed, but a large number of acquired samples are discarded while traditional image compression method is used in storage or transmission.

CS is a new method to capture and represent compressible signals at a rate significantly below the Nyquist rate. Theoretical results show that if the signal is sparse or nearly sparse in some basis, then with high probability, the measurement essentially encode the salient information in the signal, further, the unknown signal can be estimated from these compressive measurements to within a controllable mean-squared error.

Suppose any signal in  $\mathbb{R}^m$  can be expressed by linear combination of  $N \times 1$  basis vector, and suppose these basis are standardized orthogonal. Let Vector  $\{\Psi_i\}_{i=1}^N \in \mathbb{R}^N$  be a column vector to form a  $N \times N$  basis matrix  $\Psi = [\Psi_1, \Psi_2, \dots, \Psi_N]$ , So any time-domain signal  $x$  which is real, finite and discrete with one dimension can be expressed as:

$$x = \Psi\Theta = \sum_{i=1}^N \theta_i \Psi_i \quad (89.1)$$

where, vector  $\Theta$  is the coefficients of  $x$  which is represented in basis  $\Psi$  sparsely (decomposed).

$X$  and  $\Theta$  are equal expresses of same signals,  $x$  is spatial express of the signal, and  $\Theta$  is the express of signal. In  $\Psi$ -domain especially, if the non-zero number in vector  $\Theta$  is  $K$ , it's called  $k$ -sparse.

If  $\Theta$  is sorted and attenuated with power law in basis  $\Psi$ ,  $x$  is compressible.

Generally speaking, signal is not sparse itself, but when it is transformed (like wavelet transform), the coefficients can be considered sparse. For example, when a

signal is transformed by wavelet transform, we can reserve  $k$  bigger components of the coefficients, and set other  $n-k$  components to zero (because their contribution to signal reconstruction is very small), and then obtain approximate reconstructed data by inverse wavelet transform. Thus  $x$  can be considered  $K$ -sparse in wavelet basis  $\Psi$ .

Let  $\Theta$  be linear transformed by measurement matrix  $\Phi \in \mathbb{R}^{(M \times N)}$ .

Where measurement times  $M \ll N$ , measure result can obtained as following:

$$y = \Phi x = \Phi \Psi \Theta \quad (89.2)$$

where  $y = \{y_i\}_{i=1}^M$  is considered as linear projection. The dimension of  $y$  is much lower than the dimension of  $x$ . When  $x$  is reconstructed by  $y$ , it can be exactly reconstructed with a high probability from measurement results by solving the optimal problem of  $l_0$  norm.

$$\tilde{x} = \arg \min \|x\|_0 \quad s.t. \quad y = \Phi x \quad (89.3)$$

The problem of solving  $l_0$  norm is a NP-hard problem, so the problem can be changed as following:

$$\tilde{x} = \arg \min \|x\|_1 \quad s.t. \quad y = \Phi x \quad (89.4)$$

where  $l_1$ : the optimal problem with smallest norm. Algorithms like Match Pursuit, Orthogonal Match Pursuit [5], Gradient Projection, Chain Pursuit etc. are current solutions.

### 89.2.3 Compressive Sensing Algorithm Based on NSCT

When CS theorem is used in image processing, measurement matrix  $\Phi$  is formed by random Gaussian matrix which obeys  $(0, 1/N)$  distribution, or  $\pm 1$  Bernoulli matrix, namely Noiselet, etc. Image is transformed to certain domain, all of the transform coefficients are measured by  $\Phi$ , and then  $M \times N$  measurement coefficients are obtained. When recovering image, OMP algorithm can be used.

The procedure of CS algorithm based on NSCT is as following:

1. Decomposing the  $N \times N$  size image by NSCT, getting the coefficients of high-frequency and low-frequency sub-bands.
2. Selecting suitable value of  $M$  to get the measurement matrix  $\Phi$  which is  $M \times N/2$  size and Gaussian distribution, measuring the high-frequency sub-band coefficients.
3. Using OMP algorithm to reconstruct the high-frequency sub-band coefficients, and combining the low-frequency sub-band coefficients to do inverse transformation of NSCT to get the recovery image.



**Fig. 89.1** Comparing results between algorithm based on Wavelet and NSCT, (a) is original image, (b), (c) and (d) are reconstructed image based on Wavelet when  $M = 200, 150, 100$  respectively. (e) and (f) are reconstructed image based on NSCT when  $M = 100, 50$  respectively

Choosing Lena as simulation objective image, measurement matrix  $\Phi$  is random Gaussian matrix which obeys  $(0, 1/N)$  distribution, reconstruction method is OMP algorithm. High-frequency coefficients are measured by matrix  $\Phi$  but low-frequency coefficients are preserved, measurement times  $M$  is 10, 30, 50, ..., 210 respectively. On the other hand, compressed sensing algorithms based on wavelet transform is simulated as a contrast. Result shows in Fig. 89.1.

The above figure shown that very few measured times  $M$  in algorithm based on NSCT can reconstruct high quality image, but in algorithm based on wavelet more measure times are needed. In the CS algorithm based on wavelet, wavelet decomposition level has significant impact on reconstruction results, the less the decomposition level is, the worse the reconstruction results is. With the increase of the decomposition level, reconstruction effects will be improved. That's because original image could be decomposed to low-frequency sub-band and high-frequency sub-band by wavelet decomposition. High-frequency sub-band can be considered sparse, but low-frequency sub-band is the approach signal of original image under different scales, it cannot be considered sparse. When measurement matrix  $\Phi$  is multiplied by low-frequency and high-frequency coefficients together, the correlation among low-frequency approximate components coefficients will be damaged, which will deteriorate reconstruction results. When the number of wavelet decomposition level is 1, the reconstruction image is completely different from its origin. So, the wavelet decomposition level should be as large as possible. Even so, the recovered image quality is less satisfying than the CS algorithm based on NSCT.

## 89.3 Image Fusion Algorithms Combined NSCT and CS

### 89.3.1 Procedures of Image Fusion Algorithms

Using CS theory in NSCT domain, the image fusion algorithms procedures are following:

1. Decomposing the  $N \times N$  size image by NSCT, Calculate the NSCT  $j$ th layer coefficients  $C_{A,j}(x,y)$ ,  $C_{B,j}(x,y)$  of the original image A and B respectively. getting the coefficients of high-frequency and low-frequency sub-bands.
2. Selecting suitable value of  $M$  to get the measurement matrix  $\Phi$  which is  $M \times N/2$  size and Gaussian distribution, measuring the high-frequency sub-band coefficients. Preserving the low-frequency coefficients
3. Calculating fused coefficients according to some fusion rules.

In order to get better fusion effect, different fusion rules are used in low frequency and high frequency domain.

For high frequency coefficients fusion rules, using following fusion rule:

$$C_{F,j}(x, y) = \arg_{i=1, \dots, I} \max(|C_{i,j}(x, y)|) \quad (89.5)$$

For low frequency coefficients, fusion rule is follow:

$$C_{F,j}(x, y) = \frac{1}{I} \sum_{i=1}^I C_{i,j}(x, y) \quad (89.6)$$

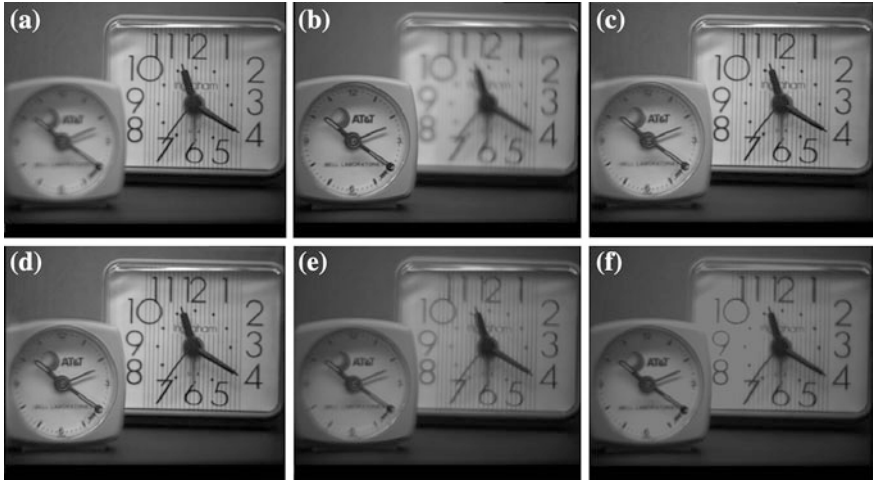
where,  $C_{F,j}(x,y)$  are the composite coefficients,  $C_{i,j}(x,y)$  are the measured NSCT coefficients,  $I$  is the total number of the source images.

4. Using OMP algorithm to reconstruct the high-frequency sub-band coefficients, and combining the low-frequency sub-band coefficients to form reconstructed NSCT coefficients  $C_{F,j}(x,y)$ .
5. Using  $C_{F,j}(x,y)$  as NSCT coefficients and do inverse transformation of NSCT to get the fused recovery image.

### 89.3.2 Experimental Simulation

In this section, two multi-focus images are selected to make fusion, wavelet domain average method, wavelet domain maximum region value method, traditional PCNN method and NSCT domain compressed sensing method are used to make comparison [14, 15]. In order to evaluate image fusion algorithms capability correctly, evaluation targets such as standard deviation, entropy and mutual information are used besides vision effect evaluation.

Figure 89.2 shows the fusion result. One can obviously find that the fused images of shift-invariant methods are clearer and more natural than the WT and



**Fig. 89.2** Multi-focus image and fusion results. **a** ClockA: Focus on *right*, **b** ClockB: Focus on *left*, **c** CS in NSCT domain, **d** PCNN, **e** Wavelet domain maximum, **f** Wavelet domain average

**Table 89.1** Objective criteria compare about several fusion algorithms

Algorithm	Standard deviation	Entropy	Mutual information
Wavelet domain average	90.80	5.04	20.12
Wavelet domain maximum	89.58	5.35	19.81
PCNN	85.39	5.69	20.05
NSCT domain compressed sensing	87.52	5.81	20.22

PCNN fused results. It is proven that NSCT domain compressed sensing methods can overcome the pseudo-Gibbs phenomena successfully and improve the quality of the fused image around edges.

Table 89.1 shows the comparing results, from it one can know, wavelet domain average method has big standard deviation, detail information is missing seriously. Wavelet domain maximum region value method is better than average method, but losing a little texture and detail information. Traditional PCNN has good performance. But NSCT domain compressed sensing method proposed in this paper has bigger mutual information, so it include more original image information.

### 89.4 Conclusions

According to the characteristic of high and low-frequency sub-band coefficients in NSCT image transform, Image fusion algorithm based on NSCT domain compressed sensing theory is proposed, it is a kind of high-efficiency fusion algorithm

fitting human vision properties. It can reserve image edge and texture well. As well as include more original image information. Compared with the some algorithms based on wavelet, the performance of the proposed algorithm is better, fewer measure times can reconstruct better recover image. But how to choose their parameters correctly need to be research deeply.

## References

1. Tseng DC, Chen YL, Liu MS (2001) Wavelet-based multi-spectral image fusion. In: Proceedings of the IEEE international conference on geoscience and remote sensing symposium (IGARSS), vol 2001(4), pp 1956–1958
2. Do MN, Vetterli M (2005) The contourlet transform: an efficient directional multiresolution image representation. *IEEE Trans Image Process* 14(12):2091–2106
3. Donoho D (2006) Compressed sensing. *IEEE Trans Inf Theor* 52(4):1289–1306
4. Candes E, Romberg J, Tao T (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inf Theor* 52(4):489–509
5. Tropp J, Gilbert AC (2007) Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans Inf Theor* 53(12):4655–4666
6. Baraniuk R (2007) A lecture on compressive sensing. *IEEE Signal Process Mag* 24(4):118–121
7. Candès EJ, Wakin MB (2008) An introduction to compressive sampling. *IEEE Signal Process Mag* 25(2):21–30
8. Candès E, Romberg J (2007) Sparsity and incoherence in compressive sampling. *Inverse Prob* 23(3):969–985
9. Donoho DL, Tsai Y (2006) Extensions of compressed sensing. *Signal Process* 86(3):533–548
10. Lu G (2007) Block compressed sensing of natural images. In: Proceedings of the international conference on digital signal processing. IEEE Press, pp 403–406
11. Zheng YA, Song JS, Zhou WM et al (2007) False color fusion for multiband SAR images based on contourlet transform. *Acta Automatica Sinica* 33(4):337–341
12. Cunha AL, Zhuo J, Do MN (2006) The nonsubsampling contourlet transform: theory, design and applications. *IEEE Trans Image Process* 15(10):1779–1793
13. Qu XB, Xie GF, Yan JW et al (2007) Image fusion algorithm based on neighbors and cousins information in nonsubsampling contourlet transform domain. In: Proceedings of the international conference on wavelet analysis and pattern recognition. IEEE, Beijing, China, pp 1797–1802
14. Li M, Cai W (2006) A region-based multi-sensor image fusion scheme using pulse coupled neural network. *Pattern Recogn Lett* 27(26):1948–1956
15. Yu RX, Zhu B, Zhang K (2008) New image fusion algorithm based on PCNN. *Opto-Electron Eng* 35(1):126–130

# Chapter 90

## The Reverse Loop Subdivision Algorithm on Approximate Minimum Error

Boning Ma, Longxing Kong, Xiaoan Tang and Gangyao Kuang

**Abstract** A new reverse Loop subdivision algorithm based on approximate minimum error is presented. At first, all isolated reverse equations in a finite grid region are found out according vertices relationship and symmetry of Loop subdivision mesh. And an integrated reverse equation is constructed using all these isolated equations by undecided parameter method. Then, the error between original grid and re-construction grid is calculated by vary original vertex. And undecided parameters are solved by using minimization error condition. To make undecided parameters independent on valence of neighbors, some approximation is applied on calculating mean error. At last, the integrated equation's computing effect on whole grid is analyzed and an equivalent computation is obtained, and a fast implementing steps is given. The experimental results show that our algorithm can get a stable reverse result for those distorted subdivision grid.

**Keywords** Loop subdivision · Reverse subdivision · Multi-scale representation

### 90.1 Introduction and Related Work

Subdivision techniques are now widely used in modeling applications. Loop subdivision is presented by Loop in 1987 [1]. It can get  $C^2$  continuity and the result grids are triangles which can be conveniently preceded by graphics hardware. This paper focuses on the reverse Loop subdivision.

The existed reverse Loop subdivision algorithms can be classified three types.

---

B. Ma (✉) · L. Kong · X. Tang · G. Kuang  
Department of Information Engineering, College of Electronic Science and Engineering,  
National University of Defense Technology, Changsha 410073, China  
e-mail: boning\_ma@sina.com



The first type is using wavelet directly for reverse Loop subdivision [2, 3]. The fine grid vertices are divided into odd vertices and even vertices. The odd vertices are predicted by the even. In this type algorithm, the coarse grid vertices are picked out directly from original grid and the result is not an optimum one.

The second type is using wavelet indirectly. These algorithms calculate coarse grid vertices from vertex-vertices and adjacent edge-vertices according to Loop subdivision scheme. In the simple solution, a coefficient is above 1 which makes algorithm unstable [4, 5], so different improved methods are issued. Bertram [4] and Li et al. [5] construct an approximate biorthogonal Loop-subdivision wavelet which adds a lifting operation on simple solution and can get stable results. Its disadvantage is that the reverse coefficients are related with the valence of computing point's neighbors.

The third type is not based on wavelet and it need not convert the reverse solution to wavelet form. Sadeghi and Samavati [6, 7] solve coarse vertices from vertex-vertices and their neighboring edge-vertices. And then, the coarse grid is adjusted according local error and surface smoothness. This method needs two times computation on whole grid. In [8], Samavati calculates the coarse vertices from vertex-vertices and their neighboring edge-vertices. They extend their idea to texture, but their algorithm is unstable.

In allusion to the problems, we present a method which is called reverse Loop subdivision algorithm on approximate minimum error (RLSAME). We derive a strict reverse Loop subdivision equation with undecided parameters. And then we solve the parameters using approximate minimum error criterion. At last, we give the fast implementation method of RLSAME. The RLSAME algorithm is stable and needs only one times computation on grid.

## 90.2 The RLSAME Algorithm

### 90.2.1 The Deduction of Reverse Formula with Undecided Parameters

Denote  $K'_l$  as a Loop subdivision grid after  $l$  subdivision steps.  $K_l$  is the edited result of  $K'_l$ , so  $K_l$  is no longer a subdivision grid.  $\tilde{K}_{l-1}$  is  $K'_l$ 's reverse subdivision grid and  $K''_l$  is subdivision result of  $\tilde{K}_{l-1}$ . The vertices of  $K'_l, K_l, \tilde{K}_{l-1}$  and  $K''_l$  are denoted respectively as  $v', v, \tilde{v}$  and  $v''$ . The vertices sets of  $K_l$  and  $K''_l$  are denoted as  $V$  and  $V''$ . The relation of all these grids is depicted in Fig. 90.1. Obviously  $K'_l$

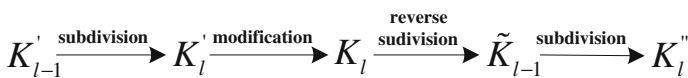
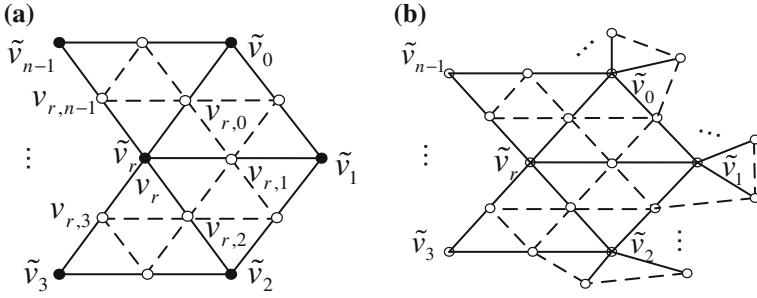


Fig. 90.1 Relationship of different grids



**Fig. 90.2** Topological relationship of loop division and reverse division. **a** the focus vertex  $v_r$  and its 2-neighborhood, **b** 3-neighborhood of the focus vertex

is not equal to  $K_l$ . Our aim is to reduce the  $K_l''$ 's deviation from  $K_l$  and at the same time the algorithm should be a strict inverse algorithm.

At first, we set there is none modification on  $K_l'$ . For vertex number is increased in subdivision process, there may be more than one equation to calculate reverse vertices. For simplicity, in this paper, we use only 2-neighborhood to find isolated reverse equations. The topological relation between  $K_l$  and  $\tilde{K}_{l-1}$  is shown in Fig. 90.2.  $v_r$  or  $\tilde{v}_r$  is our focus point.  $\tilde{K}_{l-1}$  is depicted by solid lines and solid points marked with  $\tilde{v}_i$ .  $K_l$  is depicted by all lines and points.  $v_i$  is the vertex-vertex on  $K_l$  corresponding to  $\tilde{v}_i$ .  $v_{i,j}$  is the edge-vertex on  $K_l$  corresponding to  $\tilde{v}_i$  and  $\tilde{v}_j$ . Valence of  $\tilde{v}_i$  is denoted as  $n(\tilde{v}_i)$ .

We solve Loop subdivision equation using methods in [7, 8] and obtain

$$\tilde{v}_r = 5(5 - 8n\beta)^{-1}v_r - 8\beta \cdot (5 - 8n\beta)^{-1} \sum_{j=0}^{n-1} v_{r,j} \triangleq c_r v_r + c_0 \sum_{j=0}^{n-1} v_{r,j} \tag{90.1}$$

$$\tilde{v}_i = b_r v_r + \sum_{j=0}^{n-1} (b_{(j-i)\%n} \cdot v_{r,j}), \quad i \in [0, n - 1]$$

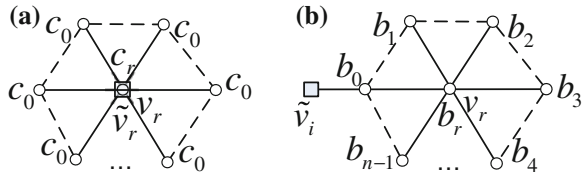
where  $n$  is the valence of  $\tilde{v}_r$ ,  $\beta = \frac{1}{n} \left( \frac{5}{8} - \left( \frac{3}{8} + \frac{1}{4} \cos \frac{2\pi}{n} \right)^2 \right)$ ,  $\alpha = 1 - n\beta$ .

Equation (90.1) can be explained using masks in Fig. 90.3, in which rectangle means pending vertex and circle means vertex on  $K_l$ . The computation of  $\tilde{v}_r$  is depicted as Fig. 90.3a which is called concentric stencil. The computation of  $\tilde{v}_i$  is depicted as Fig. 90.3b which is called eccentric stencil. We symbolize concentric stencil as  $f(\tilde{v}_r|v_r)$  and Eccentric stencil as  $g(\tilde{v}_i|v_r)$ .

We can get more equation by analyzing Fig. 90.2b. If we let  $v_0$  be the center of eccentric stencil and observe  $\tilde{v}_r$ ,  $\tilde{v}_r$  can be calculated by  $g(\tilde{v}_r|v_0)$ . In the same arguments,  $\tilde{v}_r$  can be calculated by  $g(\tilde{v}_r|v_1), g(\tilde{v}_r|v_2), \dots, g(\tilde{v}_r|v_{n-1})$ . And if we observe  $\tilde{v}_i$  using concentric stencil,  $\tilde{v}_i$  can be calculated by  $f(\tilde{v}_i|v_i)$ . So, we get

$$f(\tilde{v}_i|v_i) - g(\tilde{v}_i|v_r) = 0, \quad i \in [0, n - 1] \tag{90.2}$$

**Fig. 90.3** **a** Concentric stencil **b** Eccentric stencil



Considered  $\tilde{v}_i$ 's symmetry properties relative to  $\tilde{v}_r$ , all isolated reverse equation determined by 2-neighborhoods can be summarized as

$$\tilde{v}_r = f(\tilde{v}_r|v_r); \quad \tilde{v}_r = \frac{1}{n} \sum_{i=0}^{n-1} g(\tilde{v}_r|v_i); \quad 0 = \frac{1}{n} \sum_{i=0}^{n-1} (f(\tilde{v}_i|v_i) - g(\tilde{v}_i|v_r)) \quad (90.3)$$

We symbolize undecided parameters as  $\lambda_1$  and  $\lambda_2$ . And then we can construct integrated reverse Loop subdivision formula as

$$\tilde{v}_r = \lambda_1 \cdot f(\tilde{v}_r|v_r) + (1 - \lambda_1)n^{-1} \sum_{i=0}^{n-1} g(\tilde{v}_r|v_i) + \lambda_2 n^{-1} \sum_{i=0}^{n-1} (f(\tilde{v}_i|v_i) - g(\tilde{v}_i|v_r)) \quad (90.4)$$

### 90.2.2 Solution of Undecided Parameters

Now, we vary  $K'_i$  with  $v = v' + \Delta v'$  in Fig. 90.1. We notice that Loop subdivision schema and Eq. (90.4) are all local linear computation. If we don't consider the difference between vertex-vertex and edge-vertex, we can write  $K''_i$ 's variation  $\Delta v''$  produced by  $\Delta v'$  as.

$$\Delta v''_i = \sum_{j \in A_i} a_{i,j} \Delta v'_j \quad (90.5)$$

where  $a_{i,j}$  is the coefficient which is function of  $\lambda$ .  $A_i$ , which is related with index of  $i$ , represents indexes set of  $K''_i$ 's vertex that will affect  $\Delta v''_i$ . Then we have

$$(\Delta v''_i)^2 = \left( \sum_{j \in A_i} a_{i,j} \Delta v'_j \right)^2 = \sum_{j \in A_i} a_{i,j}^2 \Delta v_j'^2 + 2 \sum_{j \in A_i, j \neq k} a_{i,j} a_{i,k} \Delta v'_j \Delta v'_k \quad (90.6)$$

We can divide Eq. (90.6) into two parts:  $\sum_j a_{i,j}^2 \Delta v_j'^2$  and  $2 \sum_{j \neq k} a_{i,j} a_{i,k} \Delta v'_j \Delta v'_k$ .

Obviously, the first part is larger than the second part. The components in the first part are all nonnegative numbers and can not counteract each other. The second part is decided by the correlation of  $\Delta v'$  and can be neglected if  $\Delta v'$  is homogeneous distribution. So we omit the second part. Then we can compute the subdivision error of  $K''_i$  to  $K_i$  approximately as

**Table 90.1** The results of undecided parameters for common valence ( $n$ : valence)

$n$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
$\lambda_1$	-0.189	0.2508	0.4242	0.535	0.575	0.5985	0.6057	0.6066
$\lambda_2$	0.561	0.0909	-0.0955	-0.2231	-0.272	-0.3033	-0.3159	-0.3209

$$E = \|V'' - V\|^2 = \sum_i (\Delta v''_i - \Delta v'_i)^2 \approx \sum_i \sum_j (a_{ij} - b_j)^2 \Delta v_j'^2 \quad b_j = 0, 1 \quad (90.7)$$

From Eq. (90.7), we can see that the subdivision error by varying  $K'_l$ 's all vertices at the same time can be approximated by summarizing all subdivision errors by varying  $K'_l$ 's vertex one by one. This approximation means we can vary only one vertex in  $K'_l$  and use  $\min(\|V'' - V\|)$  in this situation to solve  $\lambda_i$ .

Now, vertices difference is considered. For each vertex type, we just need to calculate one  $\|V'' - V\|$  by varying one vertex in  $K'_l$ . We solve  $\lambda_1, \lambda_2$  by minimum error energy and the results for common valences are listed in Table 90.1.

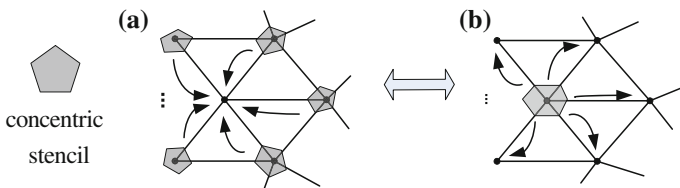
### 90.2.3 Fast Implementation of RLSAME

In Eq. (90.4), each vertex in  $\tilde{K}_{l-1}$  needs to compute  $3n + 1$  stencils (including concentric stencil and eccentric stencil) and the computation is inefficient. We can rewrite Eq. (90.4) as

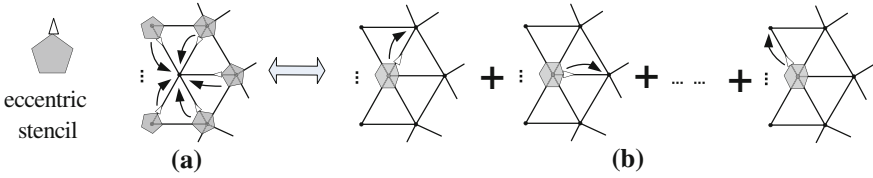
$$\tilde{v}_r = \frac{1}{n} \left[ n \cdot \lambda_1 \cdot f(\tilde{v}_r|v_r) - \lambda_2 \sum_{i=0}^{n-1} g(\tilde{v}_i|v_r) + (1 - \lambda_1) \sum_{i=0}^{n-1} g(\tilde{v}_r|v_i) + \lambda_2 \sum_{i=0}^{n-1} f(\tilde{v}_i|v_i) \right] \quad (90.8)$$

In the following analysis,  $\tilde{v}_k$  is arbitrary vertex in  $\tilde{K}_{l-1}$  and  $\tilde{v}_i, i = 0, 1, \dots, n - 1$  is  $\tilde{v}_k$ 's immediate neighbors in  $\tilde{K}_{l-1}$ . Their corresponding vertices in  $K_l$  are  $v_k$  and  $v_i$ .  $v_{k,i}$  is the edge-vertex connected with  $v_k$  and  $v_i$ .

When  $\tilde{v}_k$  is computed by Eq. (90.8), component  $\sum_{i=0}^{n-1} f(\tilde{v}_i|v_i)$  means each  $v_i$  is needed to apply concentric stencil and their results should be added to  $\tilde{v}_k$ , which is showed in Fig. 90.4a. The same process should be applied to  $\tilde{v}_i$  as  $\tilde{v}_k$  does, which means  $v_k$ 's concentric stencil result should be added to  $\tilde{v}_k$ 's immediate neighbors.



**Fig. 90.4** Equivalence effect of concentric stencil (solid points are vertex-vertices). **a** the computing process of original formula, **b** the computing process of equivalent formula



**Fig. 90.5** Equivalence effect of eccentric stencil (hollow triangle marks the position of  $b_0$ ). **a** the computing process of original formula, **b** the computing process of equivalent formula

So, we can convert the computation of Fig. 90.4a to the form of Fig. 90.4b as far as the final result of whole grid is concerned. The process in Fig. 90.4b is depicted as concentric stencil should be applied on  $v_k$  and the result is added to each  $\tilde{v}_i$  which is  $\tilde{v}_k$ 's immediate neighbor.

Component  $\sum_{i=0}^{n-1} g(\tilde{v}_k|v_i)$  means each  $v_i$  is needed to apply eccentric stencil and their results should be added to  $\tilde{v}_k$ . The coefficient  $b_0$  should be acted on  $v_{k,i}$  when eccentric stencil is used on  $v_i$ . The process is showed in Fig. 90.5a. The same process should be applied to  $\tilde{v}_i$  as  $\tilde{v}_k$  does. So, in the same argument as concentric stencil, we can convert the computation of Fig. 90.5a to the form of Fig. 90.5b. The process in Fig. 90.5b is depicted as eccentric stencil is applied on  $v_k$  with  $b_0$  acting on  $v_{k,i}$  and the result is added to  $\tilde{v}_i$  until eccentric stencil rotates one circle.

According above analysis, we can present our fast implementing method as

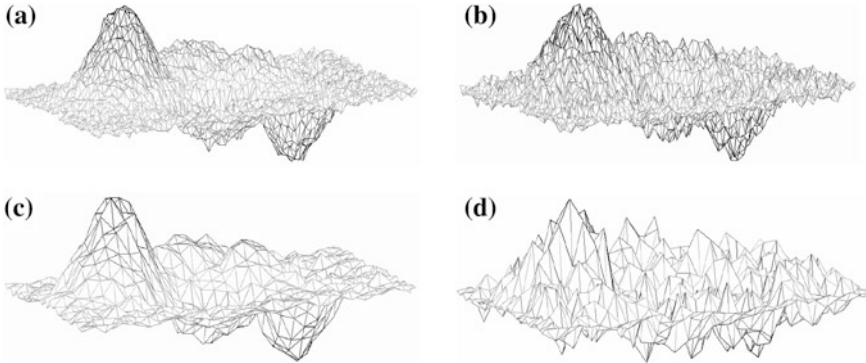
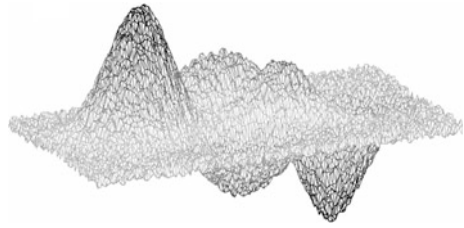
- Step 1: Allocate  $\tilde{v}$ 's storage and set its value to zero.
- Step 2: Apply concentric stencil on vertex-vertex  $v_k$  in  $K_l$  and the result is  $f(v_k)$ .
- Step 3: For  $\tilde{v}_k$  in  $\tilde{K}_{l-1}$ , compute  $\tilde{v}_k =: \tilde{v}_k + n(v_k) \cdot \lambda_{1,n(v_k)} \cdot f(v_k)$ .
- Step 4: Find all  $\tilde{v}_k$ 's immediate neighbor in  $\tilde{K}_{l-1}$  and denote the set as  $A(\tilde{v}_k)$ .
- Step 5: For each vertex  $\tilde{v}_i$  in  $A(\tilde{v}_k)$ , compute  $\tilde{v}_i =: \tilde{v}_i + \lambda_{2,n(v_i)} \cdot f(v_k)$ .
- Step 6: For each  $\tilde{v}_i$  in  $A(\tilde{v}_k)$ , compute  $g(\tilde{v}_i|v_k)$ ,  $\tilde{v}_k =: \tilde{v}_k - \lambda_{2,n(v_k)} g(\tilde{v}_i|v_k)$  and  $\tilde{v}_i =: \tilde{v}_i + (1 - \lambda_{1,n(v_i)})g(\tilde{v}_i|v_k)$ .
- Step 7: Repeat step 2–6 until all vertex-vertices in  $K_l$  are proceeded.
- Step 8: For each  $\tilde{v}_k$ , compute  $\tilde{v}_k =: \tilde{v}_k/n(v_k)$ . And the  $\tilde{v}_k$  is the final result.

In above steps, for each vertex  $v_k$ , we have one concentric stencil applied and  $n(v_k)$  eccentric stencil applied. The total applied stencil is  $n(v_k) + 1$  which is less than  $3n(v_k) + 1$  by using Eq. (90.4) directly.

### 90.3 Experiment and Discussion

The grid for experiment is showed in Fig. 90.6. The data is produced by adding Gauss noises on a smooth grid. Our algorithm belongs to non-wavelet based algorithm and the computation process is similar with [8]. Therefore we compare our algorithm with [8] in the experiment.

**Fig. 90.6** Grid for experiment



**Fig. 90.7** The result comparison of RLSAME and algorithm in reference [8]. **a** 1 level reverse results of RLSAME, **b** 1 level reverse results of [8], **c** 2 level reverse results of RLSAME, **d** 2 level reverse results of [8]

Figure 90.7 shows the reverse subdivision results on two scales. As shown, the RLSAME algorithm’s results can fit the original grid very well, while [8] get serious distorted results with reverse level rising.

The RLSAME algorithm and [8] are all strict reverse algorithms. But when original grid is far from subdivision grid, [8] can’t get similar form and this shows it is unstable. While RLSAME algorithm can get well result in this situation because RLSAME algorithm uses more fine vertices to estimate reverse vertices and its undecided parameter is solved on approximate minimum error condition.

### 90.4 Conclusion

We have presented a new reverse Loop subdivision algorithm which is abbreviated to RLSAME. Our approach exploits several reverse equations and we use these isolated equations to construct our reverse Loop equation by undecided parameters. We modify vertices’ value in original grid and then calculate the error between original grid and reconstructed grid from reverse results. We use the minimum error condition to solve undecided parameters. Some approximation is adopted when we calculate the minimum error to make undecided parameters

depend only on valence of focus vertex. To improve computation efficiency, we analyze the equivalence effect of our reverse stencils and get a fast implementation method. The fast implementation method can reduce most stencil computation for one reverse vertex.

## References

1. Loop CT (1987) Smooth subdivision surfaces based on triangles. Department of Mathematics, The University of Utah, Salt lake city
2. Ma JP, Luo XN, Ling RT et al (2007) Progressive mesh generation and its application in mobile computing. *J Image Graph* 12(2):250–255 (in Chinese)
3. Ma JP, Luo XN, Chen B et al (2009) Triangle mesh compression based on reverse subdivision for mobile terminals. *J Softw* 20(9):2607–2615 (in Chinese)
4. Bertram M (2004) Biorthogonal loop-subdivision wavelets. *Computing* 72(1–2):29–39
5. Li DG, Qin KH, Sun HQ (2004) Unlifted loop subdivision wavelets. In: Proceedings of the computer graphics and applications, 12th Pacific conference. IEEE Computer Society Washington, pp 25–33
6. Sadeghi J, Samavati FF (2009) Smooth reverse subdivision. *Comput Graph* 33(3):217–225
7. Sadeghi J, Samavati FF (2011) Smooth reverse loop and Catmull-Clark subdivision. *Graph Model* 73(5):202–217
8. Samavati FF, Pakdel HR, Smith C (2007) Reverse loop subdivision for geometry and textures. *Iran J Math Sci Inf* 2(1):21–37

# Chapter 91

## A Hybrid Algorithm Based on PBIL Algorithm and Zooming Algorithm and Its Convergence Proof

Gaopeng Wang

**Abstract** A hybrid algorithm (HA) based on population based incremental learning (PBIL) algorithm and zooming algorithm (ZA) is proposed, and its convergence is proved in this paper. In the hybrid algorithm, PBIL algorithm is employed for the evolutionary process as it can accelerate the convergence speed by a reduced time complexity, zooming algorithm is used to improve the PBIL algorithm as it can reduce search space on a large scale, and develop the convergence speed and the precision of solution obviously. The convergent analysis shows that if the population is big, and the parameters are proper, the hybrid algorithm converges to the global optimal solution.

**Keywords** Hybrid algorithm (HA) · Population based incremental learning (PBIL) · Zooming algorithm (ZA) · Convergence proof

### 91.1 Introduction

Estimation of distribution algorithm (EDA) is a well-known stochastic optimization technique, which is a combination of the statistical learn theory and random optimization algorithm [1, 2]. Population based incremental learning (PBIL) algorithm is a kind of EDA, it is based on iteratively evolving the genome of a search population by updating a probability vector, guided by the extent of class-separability demonstrated by a combination of features [3]. By virtue of its straight-forward design philosophy and implementation simplicity, PBIL has attracted the attention of many researchers, and has been used very successfully in solving a wide range of optimal problems in many disciplines [4].

---

G. Wang (✉)

National Iron and Steel Making Plant Integration Research Center, 400013 Chongqing, People's Republic of China  
e-mail: wgp20000@hotmail.com



Binary coding is used in PBIL algorithm. The convergent process of PBIL is the highest position of binary code convergence first, then followed the second high, third high etc. This order convergence process is named as domino phenomenon [5], which suffers the disadvantages of slow convergence, long time running and low precision of solutions. In order to overcome these disadvantages, a zooming algorithm (ZG) is proposed in Refs. [6, 7]. ZA can pinpoint an optimal solution at any level of desired accuracy, and it is used in various optimization problems, especially in high accuracy optimization problems [8].

The purpose of this paper is to summarize the hybrid algorithm (HA) proposed in Refs. [6, 7], and the convergence proof of the hybrid algorithm is studied.

## 91.2 PBIL Algorithm

### 91.2.1 Encoding and Decoding

Binary coding is used in the algorithm. Supposing the search space is single variable numerical interval as  $x \in [a, b]$ , the space of EDA is  $\{0, 1\}^l$ , if an individual marked as  $h = h_l h_{l-1} \cdots h_2 h_1 \in \{0, 1\}^l$ , the decoding formula is defined as

$$x = a + \left( \sum_{i=1}^l h_i 2^{i-1} \right) \frac{b-a}{2^l - 1} \quad (91.1)$$

The representing accuracy of  $x$  is  $e = \frac{b-a}{2^l - 1}$ .

### 91.2.2 Solution of Statistical Vector

The probability model is represented by a probability vector denoted as  $p(x) = (p(x_l), p(x_{l-1}), \dots, p(x_1))$ , where  $p(x)$  is the probability distribution of population,  $p(x_i) \in [0, 1]$  is the probability when the gene in position  $i$  is 1,  $1 - p(x_i)$  is the probability when the gene in position  $i$  is 0. Supposing the current generation is the  $k$ th generation, the probability vector of  $k$ th generation is  $p_k(x)$ , computing the fitness values of the  $N$  individuals and selecting the best  $T$  individuals, if  $x_k^1, x_k^2, \dots, x_k^T$  is the best  $T$  individuals, then the probability vector of  $(k + 1)$ th generation is  $p_{k+1}(x)$ , and  $p_{k+1}(x)$  has the formula as:

$$p_{k+1}(x) = (1 - \mu)p_k(x) + \mu \frac{1}{T} \sum_{j=1}^T x_k^j \quad (91.2)$$

Here  $\mu$  is the learning rate. In order to make sure the algorithm convergence to the global optimal solution, adjusting the  $p_{l+1}(x)$  to make sure  $p(x_i)$  is not too big nor too small, that is  $p(x_i) \in [\varepsilon, 1 - \varepsilon]$ . Here  $i = 1, 2, \dots, l$ ,  $\varepsilon$  is a small constant.

### 91.2.3 Population Initialization

According to the above coding method, each individual is represented as bit string composed of 0, 1 with the length of  $l$ . Initialization of population is to generate a certain number of individuals according to probability vector  $p_0(x) = (0.5, 0.5, \dots, 0.5)$ , the initialization population is defined as  $D_0$ .

### 91.2.4 Fitness Function

The fitness function is a function of variable  $x$ , expressed as

$$f = f(x) \tag{91.3}$$

### 91.2.5 Algorithm Steps

**Step 1** Confirmation of the individual length, initiation the population with  $N$  individuals according the probability vector  $p_0(x) = (0.5, 0.5, \dots, 0.5)$ .

**Step 2** Computation of the fitness of  $N$  individuals; if the result satisfies the termination conditions, iterative process finishes, if not, go to Step 3.

**Step 3** Selection of the optimal  $T = \gamma N (T < N)$  individuals as advantage community, computation probability vector of next generation with (91.2), mark the probability vector as  $p_{k+1}(x)$ .

**Step 4** Adjustment of  $p_{k+1}(x)$  to make sure  $p_{k+1}(x_i) \in [\varepsilon, 1 - \varepsilon]$ ; if  $p_{k+1}(x_i) < \varepsilon$ ,  $p_{k+1}(x_i) = \varepsilon$ ; if  $p_{k+1}(x_i) > 1 - \varepsilon$ ,  $p_{k+1}(x_i) = 1 - \varepsilon$ ,  $i = 1, 2, \dots, l, 0 < \varepsilon < 0.5$ .

**Step 5** Generation of  $N - T$  new individuals according to  $p_{k+1}(x)$ , the new population is composed of the new generated individuals and the optimal  $T = \gamma N$  individuals, back to Step 2.

## 91.3 Zooming Algorithm

Binary coding is used in PBIL algorithm, then the search space is normally represented by  $\{0, 1\}^l$ . Without loss of generality, supposing the search space is single variable numerical interval as  $[a, b]$ , while the space of PBIL is  $\{0, 1\}^l$ . If an individual marked as  $h = h_l h_{l-1} \dots h_2 h_1 \in \{0, 1\}^l$ , the decoding formula is defined as

$$x = a + \left( \sum_{i=1}^l h_i 2^{i-1} \right) \frac{b - a}{2^l - 1} \tag{91.4}$$

The representing accuracy of (91.4) is  $e_0 = \frac{b-a}{2^l-1}$ . If marks the positions of string  $h$  from right to left, the positions of  $h_1, h_2, \dots, h_l$ , are marked by position 1, position 2,  $\dots$ , and position  $l$ . The change of the number at position  $l$  means a dramatic transfer in search space.

**Definition 1:** Each string of  $\{0, 1, *\}^l$  is called a pattern of  $\{0, 1\}^l$ .

For example,  $H = 10 * 1$  is a pattern of  $\{0, 1\}^4$ . By changing the  $*$  with 1 and 0, the results as 1011 and 1001 are the samples matching with pattern  $\{0, 1\}^4$ . Patterns describe the structure comparability of binary strings.

**Definition 2:** If many individuals have same  $k$  firmly numbers from position  $l$  to its right, then they belong to pattern  $H^{lp}(s)$ , where  $s$  is the firmly number string.

For example, 10\*10, 10000, 10\*\*1 are all belong to  $H^{52}(10)$  as they both have the same binary numbers of 1, 0 in position 5 and position 4.

After  $k$  generations, if  $k > K$ , and arranging the individuals from high to low according to the fitness value, if the best  $\alpha N$  individuals at the front part belong to a same pattern marked as  $H^{lp}(s)$ , it is sure that the individuals denoted optimal solution belongs to  $H^{lp}(s)$  on the condition of  $K$  and  $\alpha$  are big enough.

For example, If  $k > K$ , and the  $\alpha N$  best individuals belong to pattern  $H^{52}(10)$ , if the  $K$  and  $\alpha$  are big enough, the optimal solution locates the certain domain of 10–11, as it is shown in Fig. 91.1.

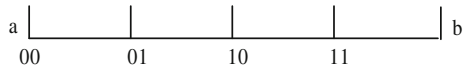
If the best  $\alpha N$  individuals belong to  $H^{lp}(s)$ , it is sure that the optimal solution belongs to  $H^{lp}(s)$ , based on the schema theorem, more and more individuals will match pattern  $H^{lp}(s)$  as the generation increased. Arrangement of the individuals from high to the low according to the fitness value, the number of individuals at the front part matching a same pattern  $H^{lp}(s)$  marked as  $M(k)$ . If the number of generation is marked as  $k$ , when  $k \rightarrow \infty$ , then  $M(k) \rightarrow N$ , here  $N$  is the number of the whole population.

If giving two parameters of  $K$  and  $\alpha$  are proper, when  $k$  exceeds  $K$  and  $M(k) > \alpha N$ , it is sure that the optimal solution locates at inferior area of  $[a, b]$ . The representing accuracy  $e$  as well as convergence speed will not been improved if the algorithm keeps on searching the interval of  $[a, b]$ , In order to overcome these disadvantages, zooming process is introduced.

According to the above analysis, if it is sure that the individual denoted optimal solution belongs to  $H^{lp}(s)$ , the searching space can be narrowed by the zooming algorithm with the follow steps.

- (1) Rearrange the individuals from the high to low according to fitness value.
- (2) Select  $\tau N$  best individuals from the population, and then deletes their  $p$  genes of  $g_i (i = l, l - 1, \dots, l - k + 1)$ , the deleted genes will been used in decoding formula.

**Fig. 91.1** The demonstration of domain of optimal solution



- (3) Shift left the remaining genes' position for  $p$ , which means that the gene in position  $i$  change to position  $i + p$ , where  $i = 1, 2, \dots, l - p$ , and then add  $p$  0 to these individual.
- (4) Generates  $(1 - \beta)N$  individuals with length of  $l$  randomly to replace the non-selected individuals. The new population is composed of the  $\beta N$  individuals and the new  $(1 - \beta)N$  individuals.

After once of this improved zooming process, the decoding formula of the new population is:

$$x = a + \left( \sum_{i=1}^p g_{l-i} \left( \frac{1}{2} \right)^{l-i+1} \right) \frac{b-a}{2^l-1} + \left( \sum_{i=1}^l h_i 2^{i-1} \right) \frac{1}{2^p} \frac{b-a}{2^l-1} \tag{91.5}$$

where  $g_i$  is the gene in the position  $i$  of the individual which has the highest fitness before the zooming process,  $h_i$  is the gene in position  $i$  of the individual in the new population after the zooming process. The accuracy of  $x$  is  $e_1 = \frac{1}{2^p} \frac{b-a}{2^l-1}$ .

After a series improved zooming process which denoted as  $H^{lp_1}(s_1), H^{lp_2}(s_2), \dots, H^{lp_j}(s_j)$ , after  $j$ th zooming process, the decoding formulas is:

$$x = a + \frac{b-a}{2^l-1} \left( \sum_{m=1}^{p_1} g_{1,l-m+1} 2^{l-m} + \sum_{m=1}^{p_2} g_{2,l-m+1} 2^{l-q_1-m} + \dots + \sum_{m=1}^{p_j} g_{j,l-m+1} 2^{l-\sum_{i=1}^{j-1} p_j-m} \right) + \left( \sum_{m=1}^l h_{j,l-m} 2^{m-1} \right) \cdot \frac{b-a}{2^{\sum_{m=1}^j p_j} (2^l-1)} \tag{91.6}$$

where  $g_{j,l-i+1} (i = 1, 2, \dots, k_j)$  denotes the  $k_i$  same genes in the  $j$ th zooming process;  $h_i^{(j)} (i = 1, 2, \dots, l)$  are the gene of position  $i$  of the new individuals after  $j$ th zooming process. The accuracy of  $x$  is  $e_j = \frac{1}{2^{\sum_{i=1}^j p_j}} \frac{b-a}{2^l-1}$ .

### 91.4 Hybrid Algorithm Based on PBIL Algorithm and ZA

The hybrid algorithm has two actions, one is the population evolution, and the other is the zooming process. According to the PBIL algorithm and the improved zooming process described above, the hybrid algorithm has the flowchart as shown in Fig. 91.2.

In the flowchart, the parameters of  $K$  and  $\alpha$  are set to guarantee that the optimal solution and the best individual are belong to a same pattern. If the value of  $K$  and  $\alpha$  are too big, the convergence speed is slow; If they are too small, the probability of the optimal solution and the best individuals belong to the same pattern is small. Pattern  $H^{lp}(s)$  is depends on the  $\alpha N$  individuals which have the higher fitness value

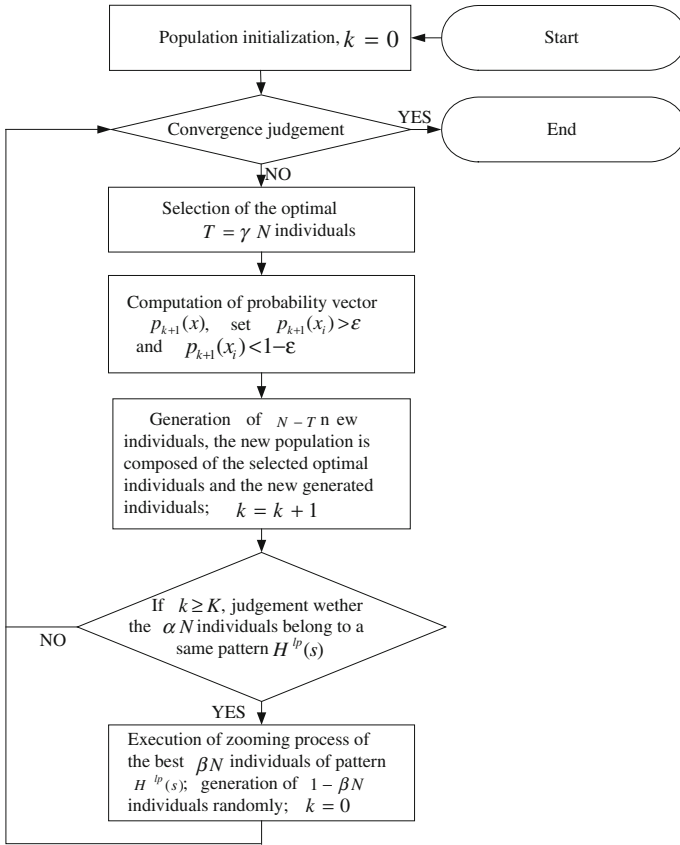


Fig. 91.2 Flowchart of the hybrid algorithm

in the population.  $\gamma$  is set for preserving the important information of the former population. If  $\gamma$  is too big, the diversity of the new population is bad. If  $\gamma$  is too small, the new population will lose the information of the former population.  $\mu$  is the learning rate of the PBIL algorithm, and  $\epsilon$  is set to make sure the PBIL algorithm convergence to the global optimal solution.

For multi-dimensional optimal problems, the fitness function are denoted as  $f(x) = f(x_1, x_2, \dots, x_t, \dots, x_n)$ , here  $x = (x_1, x_2, \dots, x_t, \dots, x_n)$ ,  $x \in D$ . If  $x_t \in [a_t, b_t]$  are coded as  $h^i (h^i = h_i^1 h_{i-1}^1 \dots h_2^i h_1^i)$ , the solution can be denoted as  $h = \underbrace{h_i^1 h_{i-1}^1 \dots h_2^1 h_1^1}_{h^1} \underbrace{h_i^2 h_{i-1}^2 \dots h_2^2 h_1^2}_{h^2} \dots \underbrace{h_i^n h_{i-1}^n \dots h_2^n h_1^n}_{h^n}$ , then the improved zoom-

ing process are executed for each  $x_t$  separately.

References [6] and [7] give many examples of using this hybrid algorithm to solve optimization problems. The results show that the hybrid algorithm is effective.

### 91.5 Convergent Analysis of Hybrid Algorithm

The optimization problem is expressed as:

$$\max f(x), \quad x \in W \tag{91.7}$$

where  $x = (x_1, \dots, x_t, \dots, x_n)$ ,  $W \in R^n$  is a nonempty bounded closed set.  $f(x)$  is continuous non-negative objective function. There must have  $x^* = (x_1^*, \dots, x_t^*, \dots, x_n^*)$  makes  $f(x) \leq f(x^*)$  for any  $x(x \in W)$ , thus  $x^*$  is named as global optimal solution, and  $G^* = f(x^*)$  is named as global maximum. If  $C < G^*$ , then  $H = \{x|x \in D, f(x) > C\}$  is non-empty set.

#### 91.5.1 Convergent Analysis of PBIL Algorithm

Supposing  $Pop(k)$  is the population of step  $k$ ,  $Pop^S(k)$  is selected from  $Pop(k)$  for founding probability model. If  $Pop(k)$  belongs to probability model  $p(x, k)$ ,  $Pop^S(k)$  belongs to probability model  $p^S(x, k)$ , if the population of  $Pop(k)$  and  $Pop^S(k)$  is big enough, the probability density functions statistic from  $Pop(k)$  and  $Pop^S(k)$  convergence to  $p(x, k)$  and  $p^S(x, k)$  separately. Therefore, if the population is big enough,  $p(x, k)$  can be used to represent the population of step  $k$ , and  $p^S(x, k)$  can be used to represent the dominate population selected from  $Pop(k)$  which is used for founding probability model. The principle of estimation of distributed algorithm can be represented as the fellow steps:

- Step 1** Dominate population  $p^S(x, k)$  is selected from population  $p(x, k)$ ;
- Step 2**  $p(x, k + 1)$  as population of next step is sampled from dominate population  $p^S(x, k)$ .

If defining

$$E(k) = \int_D f(x)p(x, k)d_x \tag{91.8}$$

then  $E(k)$  is the average fitness function. Given

$$\lim_{k \rightarrow \infty} E(k) = G^* \tag{91.9}$$

If satisfying (91.9), the estimation of distributed algorithm is convergence. That is mean all the individual of the population is the global optimal solution when the algorithm convergence.

According to the evolution process above, at step  $k$ , the selection process of PBIL algorithm is to range all the individuals at their fitness from high to low, and selecting optimal individuals as dominate population. If the selecting probability is  $\gamma$ , the selected probability model is [9]:

$$p^S(x, k) = \begin{cases} \frac{p(x,k)}{\gamma}, f(x) \geq \beta(k) \\ 0, f(x) < \beta(k) \end{cases} \tag{91.10}$$

where

$$\gamma = \int_{f(x) \geq \beta(k)} p(x, k) dx \tag{91.11}$$

Equation (91.11) means that individuals are not selected into the dominate population if their fitnesses smaller than  $\beta(k)$ , and  $\beta(k)$  is the individual of smallest fitness when the selection probability is  $\gamma$ .

PBIL algorithm selects dominate population  $Pop^S(k)$ , founds probability model from  $Pop^S(k)$ , and generates new next population from the new probability model.

Because when  $f(x) < \beta(k)$ ,  $p(x, k + 1) = 0$ , then

$$\int_{f(x) \geq \beta(k)} p(x, k + 1) dx = \int_D p(x, k + 1) dx = 1 \tag{91.12}$$

From Eq. (91.11)

$$\int_{f(x) \geq \beta(k+1)} p(x, k + 1) dx = \gamma(k + 1) < 1 \tag{91.13}$$

Comparison of Eqs. (91.12) and (91.13), it obtained that

$$\beta(k) < \beta(k + 1), k = 1, 2, \dots \tag{91.14}$$

Then there exist  $\lim_{k \rightarrow \infty} \beta(k)$ , supposing

$$\beta = \lim_{k \rightarrow \infty} \beta(k) \tag{91.15}$$

Assume that  $\beta < G^*$ , therefore

$$p(x, k) = p(x, 0) \prod_{i=0}^{k-1} [\gamma]^{-1} \geq \gamma^{-k} p(x, 0) \tag{91.16}$$

whenever  $f(x) > \beta$ . Noting that  $p(x, 0) > 0$  for any  $x \in D$ , it has

$$\lim_{k \rightarrow \infty} p(x, k) = +\infty \tag{91.17}$$

for all  $x$  with  $f(x) > \beta$ . Make  $S = \{x | x \in D, f(x) > \beta\}$ , since  $\beta < G^*$ , the Borel measure of  $S$  is positive, by Fatou's lemma it obtain

$$\lim_{k \rightarrow \infty} \int_S p(x, k) dx = +\infty \tag{91.18}$$

which contradicts the fact that  $p(x, k)$  is a probability density function. Therefore, it obtain

$$\lim_{k \rightarrow \infty} \beta(k) = G^* \tag{91.19}$$

Since

$$E(k) = \int_D f(x)p(x, k)d_x \geq \int_D \beta(k)p(x, k)d_x = \beta(k) \tag{91.20}$$

it have

$$\lim_{k \rightarrow \infty} E(k) = G^* \tag{91.21}$$

The PBIL algorithm is convergent.

### 91.5.2 Convergent Analysis of Hybrid Algorithm

In the hybrid algorithm, the PBIL algorithm is convergent. If the population is big enough, and the parameters of  $K$  and  $\alpha$  is appropriate, after one zooming process, the optimal solution is still in the searching space.

Since the zooming algorithm decreases the searching space of each variable independently, to help describe, only consider one variable.

The decoding formula of indigent variable  $x_t$  after  $j$  times zooming process is:

$$\begin{aligned} x_t^{(j)} = & a + \frac{b-a}{2^l-1} \left( \sum_{i=1}^{p_1} g_{1,l-i+1} 2^{z-i} + \sum_{i=1}^{p_2} g_{2,l-i+1} 2^{z-p_1-i} + \dots + \sum_{i=1}^{p_j} g_{j,l-i+1} 2^{z-\sum_{i=1}^{j-1} p_i-i} \right) \\ & + \left( \sum_{i=1}^{l_j} h_{l-z+i}^{(j)} 2^{i-1} \right) \cdot \frac{b-a}{2^{\sum_{i=1}^j p_i} (2^l-1)} = g_1^{(j)} + g_2^{(j)} \end{aligned} \tag{91.22}$$

where  $g_1^{(j)} (j = 1, 2, \dots)$  is a monotonically increasing sequence. Here to prove it is bounded.

Since



$$\begin{aligned}
 g_1^{(j)} &= a + \frac{b-a}{2^z-1} \left( \sum_{i=1}^{p_1} g_{1,l-i+1} 2^{z-i} + \sum_{i=1}^{p_2} g_{2,l-i+1} 2^{z-p_1-i} + \dots + \sum_{i=1}^{p_j} g_{j,l-i+1} 2^{z-\sum_{i=1}^{j-1} p_i-i} \right) \\
 &\leq a + \frac{b-a}{2^z-1} \left\{ \frac{2^{z-1} \left[ 1 - \left(\frac{1}{2}\right)^{p_1} \right]}{1 - \frac{1}{2}} + \frac{2^{z-p_1-1} \left[ 1 - \left(\frac{1}{2}\right)^{p_2} \right]}{1 - \frac{1}{2}} + \dots + \frac{2^{z-\sum_{i=1}^{j-1} p_i-1} \left[ 1 - \left(\frac{1}{2}\right)^{p_j} \right]}{1 - \frac{1}{2}} \right\} \\
 &= a + \frac{b-a}{2^z-1} \left\{ 2^z \left[ 1 - \left(\frac{1}{2}\right)^{p_1} \right] + 2^{z-p_1} \left[ 1 - \left(\frac{1}{2}\right)^{p_2} \right] + \dots + 2^{z-\sum_{i=1}^{j-1} p_i} \left[ 1 - \left(\frac{1}{2}\right)^{p_j} \right] \right\} \\
 &= a + \frac{b-a}{1-2^{-z}} \left\{ \left[ 1 - \left(\frac{1}{2}\right)^{p_1} \right] + 2^{-p_1} \left[ 1 - \left(\frac{1}{2}\right)^{p_2} \right] + \dots + 2^{-\sum_{i=1}^{j-1} p_i} \left[ 1 - \left(\frac{1}{2}\right)^{p_j} \right] \right\} \\
 &= a + \frac{b-a}{1-2^{-z}} \left( 1 - 2^{-\sum_{i=1}^j p_i} \right)
 \end{aligned}$$

$g_1^{(j)} \leq a + \frac{b-a}{1-2^{-z}}$  when  $j \rightarrow \infty$ ,  $g_1^{(j)} (j = 1, 2, \dots)$  is an increasing and bounded series, then it convergent and its limit exist [10]. Meanwhile, when  $j \rightarrow \infty$ , the limit of  $g_2^{(j)}$  exists, so the limit of  $x_t^{(j)}$  exists.

Denote

$$\lim_{j \rightarrow \infty} x_t^{(j)} = x_t^0 \tag{91.23}$$

Assuming  $x_t^0 \neq x_t^*$ , where  $x_t^*$  is the optimal solution of  $x_t$ . Since  $x_t^*$  and  $x_t^0$  belong to a same searching space, there must has  $d \in [g, h] (g \neq h)$  and  $x_t^* \in d, x_t^0 \in d$ .

According the zooming algorithm, after one zooming process, searching interval is reduced to  $2^{-p_1}$  times of original, after  $j$  times zooming process, searching interval

of  $x_0$  reduces to  $2^{-\sum_{i=1}^j p_i}$  of original. When  $j \rightarrow \infty$ , the length of search interval is close to 0, which is contradiction with the existence of  $d$ . Therefore,  $x_t^0 = x_t^*$ . That means  $x$  is convergent to  $x^*$ , the hybrid algorithm is convergent.

### 91.6 Conclusion

This paper proposes a hybrid algorithm based on the PBIL algorithm and zooming algorithm. The zooming algorithm improves the convergence speed and the precision of the solution of PBIL algorithm. On the condition of population is big enough, and the parameters of zooming algorithm are proper, it is proved that the hybrid algorithm can converges to the global optimal solution.

## References

1. Shim VA, Tan KC, Cheong CY (2012) A hybrid estimation of distribution algorithm with decomposition for solving the multiobjective multiple traveling salesman problem. *IEEE Trans Syst Man Cybern Part C Appl Rev* 12(5):682–691
2. Armananzas R, Saeys Y, Inzz I (2011) Peakbin selection in mass spectrometry data using a consensus approach with estimation of distribution algorithms. *IEEE/ACM Trans Comput Biol Bioinf* 8(3):760–774
3. Ho SL, Yang SY, Fu WN (2011) A population-based incremental learning vector algorithm for multiobjective optimal designs. *IEEE Trans Magn* 47(5):1306–1309
4. Chen XM, Lei G, Yang GY (2012) An improved population-based incremental learning method for objects buried in planar layered media. *Ann Nucl Energy* 48(2):1027–1030
5. Chen TS, Tang K, Chen GL (2007) On the analysis of average time complexity of estimation of distribution algorithms. In: *IEEE congress on evolutionary computation*, pp 453–450
6. Wang GP, Dou LH, Chen J (2009) A hybrid algorithm based on PBIL algorithm and zooming algorithm. *Pattern Recogn Artif Intell* 22(1):1–5 (in Chinese)
7. Dou LH, Wang GP, Chen J (2009) A hybrid algorithm for computing cannonball dispersion evenness. *Control Theory Appl* 26(6):624–628 (in Chinese)
8. Kown YD, Kown HW, Cho SW (2006) Convergence rate of the successive zooming genetic algorithm using optimal control parameters. *WSEAS Trans Comput* 5(6):1200–1207
9. Zhang QF, Mühlenbein H (2004) On the convergence of a class of estimation of distribution algorithms. *IEEE Trans Evol Comput* 8(2):127–136
10. Wang JJ, Li GM (2007) *Higher mathematics*. Tsinghua University Press, Beijing (in Chinese)

# Chapter 92

## Image Intensity Correction Under Illumination Changes Based on Mixture Gaussian Model

Yanxiang Han, Zhisheng Zhang, Lei Zhang, Ping Chen and Fei Hao

**Abstract** Illumination changes in video sequences result in a drastic increase in the number of falsely detected change regions and make change detection unreliable. In this paper, we propose a novel approach for intensity correction under illumination variation. A mixture Gaussian model consisting of two density components associating with two classes is used. Based on Expectation–maximization algorithm, the statistical parameter estimations are performed. Under the assumption of Gaussian distribution for stationary pixels, the global intensity factor can be calculated for image intensity correction. Finally, two experiments are carried out to verify the proposed method.

**Keywords** Intensity correction · EM algorithm · Parameter estimation

### 92.1 Introduction

All the time, dynamic sequence is widely used in remote sensing, medical imaging, optic flow and object tracking [1–3]. Dynamic sequence has been proved very successful in a well-constrained environment, where there is full control over lighting conditions. However, illumination changes of scene usually occur due to lights switching on or off, clouds moving in front of the sun. As a result, illumination variations give rise to image intensity changes over time, which result in a drastic increase in the number of falsely detected change regions and makes change detection unreliable under changing illumination conditions.

An approach [4], was proposed to remove disturbing temporal fluctuations of image intensity by equalizing local frame means and variances in the temporal

---

Y. Han · Z. Zhang (✉) · L. Zhang · P. Chen · F. Hao  
School of Mechanical Engineering, Southeast University, Nanjing, 211189 Jiangsu, China  
e-mail: oldbc@seu.edu.cn

scene based on compensating for camera pan followed by local motion detection. However, this is not the case in most practical situations, and the variance need to be estimated in other applications. Withagen et al. [5] proposed generally applicable algorithms to correct for global difference in image intensity between images recorded with a static or slowly moving camera based on intensity-quotient estimation. Subsequently, Sayed et al. [6] presented an efficient algorithm for high definition video surveillance applications by using an apparent gain factor to correct global intensity changes and by using local mean and standard deviation to correct local intensity changes. Although the global correction method performs well when illumination changes globally, the process of outlier removal is necessary to be carried out based on the choice of optimal value of non-stationary factor which may change for different scene.

In this paper, we propose an intensity correction method based on a mixture Gaussian model. It is assumed that the distribution of intensity value in the ratio image can be modeled as a mixture Gaussian model consisting of two density components. The estimations for statistical parameters relating to stationary and non-stationary pixels are performed based on Expectation–maximization algorithm.

## 92.2 Proposed Algorithm

In this section, we present an efficient estimation for intensity factor of illumination changes based on a simplified model of a CCD camera [7]. We firstly introduce a model of intensity correction [5]. Then a mixture Gaussian distribution model in the ratio image is built for the classifications of stationary and non-stationary pixels. Moreover, parameter estimations relating to two classes are performed based on EM algorithm.

### 92.2.1 Model for Intensity Correction

Based on a simplified CCD model with gamma correction, Withagen et al. [5] proposed a correction method for global changes in image intensity under illumination variations, where the intensity difference can be modeled as follows:

$$i_{t,corrected} = \frac{i_t}{a} \quad (92.1)$$

where  $i_t$  is the current image intensity,  $i_{t,corrected}$  and  $i_r$  have equal global intensity.  $a$  is the intensity factor of illumination change, which can be estimated by minimizing the criterion,

$$L = \sum_{s \in S} \omega_s^2 (i_{t,s} - a i_{r,s})^2 \quad (92.2)$$

for all pixels  $s$  in the set  $S$  in the least squares sense [5].

In the process, outlier removal was performed for non-stationary pixels based on statistics with the assumption that the majority of pixels depicted the same scene in both images. Here, the coefficient for outlier removal  $T_{outlier}$  as decided by using empirical strategies or manual trial-and-error procedures, which may affect both the accuracy of the estimation for the intensity factor.

### 92.2.2 Basic Principle for Intensity Factor Estimation

Consider two images,  $X_c = \{x_c(i,j) | 1 \leq i \leq M, 1 \leq j \leq N\}$  and  $X_b = \{x_b(i,j) | 1 \leq i \leq M, 1 \leq j \leq N\}$ . The main objective is to discriminate between the stationary pixels and non-stationary pixels from the ratio image  $X_r = \{x_r(i,j) | 1 \leq i \leq M, 1 \leq j \leq N\}$  which can be computed as intensity ratio of corresponding pixels between the current and the reference image, i.e.,

$$x_r(i,j) = \frac{x_c(i,j)}{x_b(i,j)} \quad (92.3)$$

where  $x_c(i,j)$  and  $x_b(i,j)$  is the current and reference image intensity at the location  $(x,y)$ , respectively. For all the pixels, including stationary and non-stationary pixels, the method assumes the probability density function  $p(x_r)$  computed on the pixel values in the ratio image  $X_r$  can be modeled as a mixture density distribution consisting of two density components associated with two classes  $\omega_s$  and  $\omega_n$ , respectively, i.e.,

$$p(x_r) = p(x_r|\omega_s)P(\omega_s) + p(x_r|\omega_n)P(\omega_n) \quad (92.4)$$

where  $x_r$  is the intensity value in the ratio image  $X_r$ ;  $\omega_s$  and  $\omega_n$  denote the sets of “stationary” and “non-stationary” pixels, respectively;  $p(x_r|\omega_s)$  and  $p(x_r|\omega_n)$  are posteriori probability density functions, and  $P(\omega_s)$  and  $P(\omega_n)$  are priori probabilities of the class,  $\omega_s$  and  $\omega_n$ , respectively. The estimations of  $p(x_r|\omega_s)$ ,  $p(x_r|\omega_n)$ ,  $P(\omega_s)$ ,  $P(\omega_n)$  can be performed by using EM algorithm.

### 92.3 Estimations for Intensity Factor

In this section, this goal is to estimate the set of statistical parameters  $\theta = \{u_s, u_n, \sigma_s^2, \sigma_n^2, P(\omega_s), P(\omega_n)\}$  by maximizing the log likelihood  $L(\theta; X_r)$  of the two-components mixture defined in (92.4). A possible choice for likelihood estimations of mixture parameters is to use EM algorithm [8], which performs well

for incomplete data problems. The EM algorithm is consisted of an expectation step and a maximization step, which are iterated until convergence is reached.

**E-step:**

This step is assumed that the observations  $X_r$  in the ratio image are incomplete. With respect to the previous estimates of the parameters  $\theta^t$ , the current priori probability for the class,  $\omega_s$  and  $\omega_n$ , can be computed as follows,

$$P^{t+1}(\omega_s) = \frac{\sum_{x(i,j) \in X_r} \frac{P^t(\omega_s) p^t(x(i,j)|\omega_s)}{p^t(x(i,j))}}{MN} \tag{92.5}$$

**M-step:**

In this step, the parameters which characterize the density functions  $p(x_r|\omega_s)$  and  $p(x_r|\omega_n)$  are determined according to the estimations of prior probability. This corresponds to estimating  $u_s, u_n, \sigma_s^2, \sigma_n^2$  for the two components of GMM with respect to,

$$\mu_s^{t+1} = \frac{\sum_{x(i,j) \in X_r} \frac{P^t(\omega_s) p^t(x(i,j)|\omega_s)}{p^t(x(i,j))} x(i,j)}{\sum_{x(i,j) \in X_r} \frac{P^t(\omega_s) p^t(x(i,j)|\omega_s)}{p^t(x(i,j))}} \tag{92.6}$$

$$(\sigma_s^2)^{t+1} = \frac{\sum_{x(i,j) \in X_r} \frac{P^t(\omega_s) p^t(x(i,j)|\omega_s)}{p^t(x(i,j))} [x(i,j) - \mu_s^t]^2}{\sum_{x(i,j) \in X_r} \frac{P^t(\omega_s) p^t(x(i,j)|\omega_s)}{p^t(x(i,j))}} \tag{92.7}$$

where the superscripts  $t$  and  $t + 1$  represent the current and next iteration indices, respectively.  $\mu_s$  and  $\sigma_s$  denote the mean and variance of the density function associated with the class  $\omega_s$ . The estimations for priori probabilities concerning  $\omega_s$  are calculated using the iterative variables  $P^{t+1}(\omega_s)$  and  $P^{t+1}(\omega_n)$ , respectively. The Gaussian model parameters associating with the class  $\omega_s$  are estimated using the iterative process. Analogous equations are used for the class  $\omega_n$ .

Considering the effect of noise in the ratio image, model parameters  $\mu_s$  and  $\sigma_s^2$ , can be considered as statistics of intensity factor under illumination changes. Since the values of stationary pixels neighboring the mean  $\mu_s$  have same contributions to the value of the intensity factor, it can be computed according to the following equation,

$$a = \frac{1}{|S|} \sum_{x_r(i,j) \in S} x_r(i,j) \tag{92.8}$$

where  $|S|$  denotes the number of pixels in the set of pixels  $S = \{x_r(i,j) | u_s - 0.5\sigma_s \leq x_r(i,j) \leq u_s + 0.5\sigma_s\}$ . Here,  $u_s$  and  $\sigma_s$  denote the estimations for Gaussian distribution model parameters associated with the stationary class  $\omega_s$ .

### 92.4 Experiment

In this paper, the first image of a sequence is selected as the reference image for the convenience of computation. The distorted images are generated by using synthetically global change factors applied to original images. For each frame in the test sequence, the factors for global intensity changes are obtained with the range from 0.75 to 1.25, which are randomly selected. Then original images can be distorted by using the following equation [6],

$$I(i, j) = Y(i, j)dp \tag{92.9}$$

where  $Y(i, j)$  is the intensity in the original image at location  $(i, j)$ ,  $I(i, j)$  is the distorted intensity at location  $(i, j)$  and  $dp$  is the illumination change factor.

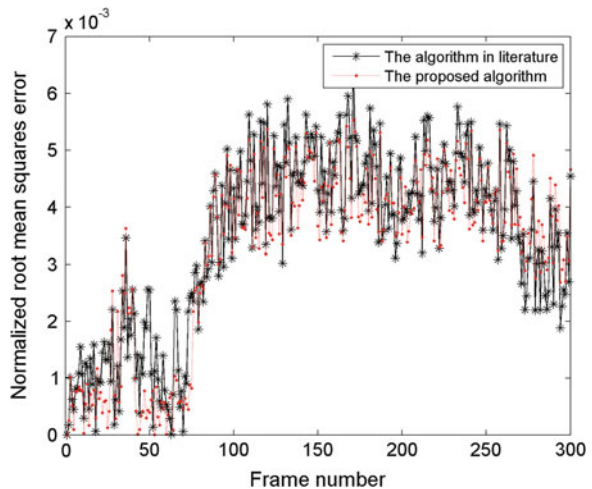
For the accuracy evaluation, we use the normalized root mean squares error between the image intensity corrected by the proposed method and the intensity in the original image for all pixels of each frame,

$$e_{accuracy} = \frac{\sqrt{\frac{1}{|S|} \sum_{i(x,y) \in S} (i_{corrected}(i, j) - i_{original}(i, j))^2}}{255} \tag{92.10}$$

where  $i_{corrected}(i, j)$  represents the image intensity at location  $(i, j)$  in the corrected image with the proposed method, and  $i_{original}(i, j)$  denotes the image intensity at location  $(i, j)$  in the original image  $|S|$  is the number of pixels in the frame.

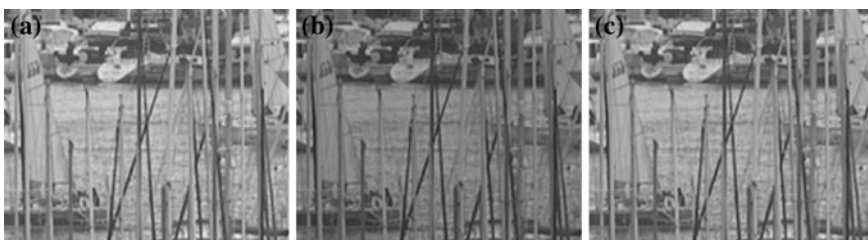
The analysis of the normalized root mean squares error for ‘HALL’ sequence is shown in Fig. 92.1. In particular, the mean of normalized root mean squares errors with the algorithm [5] and the proposed method in this paper is 0.32 and 0.34 %, respectively. Samples of the original, distorted, and corrected image are shown in Fig. 92.2.

**Fig. 92.1** The analysis for the normalized root mean squares errors for ‘HALL’





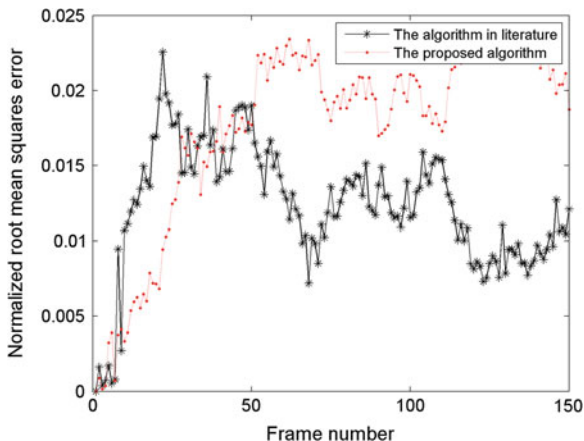
**Fig. 92.2** Examples of the original, distorted, and corrected image from ‘HALL’ sequence. **a** original image. **b** distorted image. **c** corrected image



**Fig. 92.3** Examples of the original, distorted, and corrected image from ‘HARBOUR’ sequence. **a** original image. **b** distorted image. **c** corrected image

Samples of the original, distorted, and corrected image are shown in Fig. 92.3 for the sequence ‘HARBOUR’, in which change regions between the current image and the reference image are much larger than that in the sequence of ‘HALL’. The results of normalized root mean squares errors are shown in Fig. 92.4 with the average error 1.23 % while the average error is 1.76 % provided by the literature [5].

**Fig. 92.4** The analysis for the normalized root mean squares errors for ‘HARBOUR’ sequence





It can be seen that the accuracy between the algorithm [5] and the proposed method is very close for the experiment of ‘HALL’ sequence, while the accuracy is improved for the sequence of ‘HARBOUR’. From the analysis of image sequences, it may imply that there are more complex motion and change regions in the latter sequence. In that situation, the mixture Gaussian model is more efficient to discriminate between stationary and non-stationary pixels. Although the absolute average error increases in the latter experiment due to existence of the complex motion, the relative error between the algorithm [5] and the proposed method decreases. It implies that the proposed method is more robust to complex dynamic scene for intensity correction under illumination changes.

## 92.5 Conclusions

In this paper, image intensity correction is performed by using the global intensity factor. It is assumed that the probability density function computed on the pixel values in the ratio image can be modeled as a mixture Gaussian distribution. The estimations of Gaussian model parameters can be performed by EM algorithm., and the intensity factor is obtained for image intensity correction under illumination variations. Results of experiments show that the proposed method is more robust for intensity correction under situations of the existence of more complex motions in the image sequence.

**Acknowledgments** This work was supported by National Nature Science Foundation of China (NSFC) under 50805023, the Special Fund of Jiangsu Province for the Transformation of Scientific and Technological Achievements under BA2010093 and the Hexa-type Elites Peak Program of Jiangsu Province under 2008144, the key construction laboratory of Jiangsu Province for detection and controlling of large engineering equipment.

## References

1. Lillestr R (1972) Techniques for change detection. *IEEE Trans Comput* 21(7):654–659
2. Townshend JRG, Justice CO, Gurney C, McManus J (1992) The impact of misregistration on change detection. *IEEE Trans Geosci Remote Sens* 30(5):1054–1060
3. Molnar J, Chetverikov D, Fazekas S (2010) Illumination-robust variational optical flow using cross-correlation. *Comput Vis Image Underst* 114(10):1104–1114
4. van Roosmalen PMB, Lagendijk RL, Biemond J (1999) Correction of intensity flicker in old film sequences. *IEEE Trans Circuits Syst Video Technol* 9(7):1013–1019
5. Withagen PJ, Schutte K, Groen FCA (2010) Global Intensity Correction in Dynamic Scenes. *Int J Comput Vision* 86(1):33–47
6. Sayed MS (2011) An efficient intensity correction algorithm for high definition video surveillance applications. *IEEE Trans Circuits Syst Video Technol* 21(11):1622–1630
7. Withagen PJ, Groen FCA, Schutte K (2007) CCD color camera characterization for image measurements. *IEEE Trans Instrum* 56(1):199–203
8. Dempster BAP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc B* 39(1):1–38

# Chapter 93

## Maneuver Target Detection Method with Iterative Endpoint Fitting Assisted

Zhangsong Shi and Zhonghong Wu

**Abstract** A target maneuvering detection method with iterative endpoint fitting assisted is derived and presented. With the similarity between image curve fitting and maneuvering target tracking has been thought of in the method, when the traditional target maneuvering detection method was used, the target maneuver information could be abstracted with the principle of iterative endpoint fitting assisted, and the maneuver start point would be detected more accurately by the information feed backed to the detection process. The effectiveness of the method has been verified by numerical simulation.

**Keywords** Point · Iterative endpoint fitting · Target maneuvering detection · Maneuver target tracking

### 93.1 Introduction

Maneuvering target tracking has received much attention in recent years due to its military value, which mainly includes single-mode method based on the decision-making and multi-mode method. As for single-mode method, the process is mainly composed of data preprocessing, target maneuver model, maneuver detection, filtering and prediction, etc. Maneuver detection is one of the important component modules. Despite the tracking performance of multi-mode method is better compared with single-mode method, the applicable scope of this method received limit as the design process is complex and the calculated resource consumption is tremendous [1, 2]. However, studies have shown that with a good maneuver detection method single-mode method can get a similar tracking result to multi-mode method,

---

Z. Shi · Z. Wu (✉)

Department of Command and Control, College of Electronic Engineering, Naval University of Engineering, Jiefang Dadao Road 717, Wuhan 430033 Hubei, China  
e-mail: yizhousan@163.com

in view of this fact realization target maneuver detection timely and accurate become the key of single-mode method [3, 4].

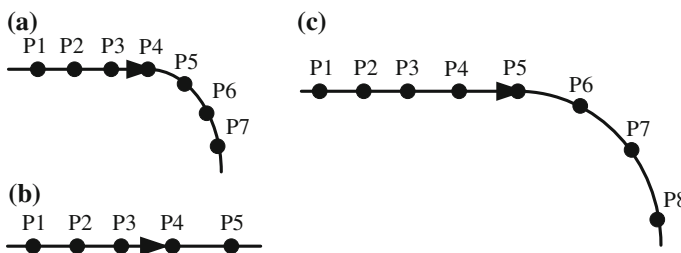
The main maneuvering detection method is sliding window detection and its improved methods [5, 6]. Target maneuver detection has been system discussed in these methods, and properties of typical maneuver detection methods have been analyzed in [7], target maneuver detection research has been helped by all these studies.

In fact, the process of maneuvering target tracking is similar to the process of image curve fitting, on account of this fact, an maneuvering target detection method with iterative endpoint fitting assisted has been presented in this paper, as a result of this method used, target maneuvering information such as start maneuvering point and start maneuvering time could be abstracted, feed these information into the process of traditional maneuvering target detection method, the details of target maneuvering could be understood better, therefore target has been tracked more accurate.

### 93.2 Algorithm Principles

Although there are many target maneuvering styles, all styles are consisted of the following two basic modes, direction changed and velocity changed. Therefore the target maneuver styles can be divided into three types that direction changed type, velocity changed type and direction changed with velocity changed type, etc., like shown in Fig. 93.1.

When a maneuver target is tracked, the difference between observed value and predicted value will increase gradually, and after a certain data processing cycle filter divergence, then target maneuver can be determined. As shown in Fig. 93.2, along with the target maneuver, the difference between observed value (point P5, point P6, point P7) and predicted value (point P5', point P6', point P7') became more and more large, and then set a threshold value; when the difference is greater than the threshold value, the target maneuver is determined.



**Fig. 93.1** Three modes of target maneuvering. **a** Direction changed maneuvering **b** Velocity changed maneuvering **c** Both direction and velocity changed maneuvering

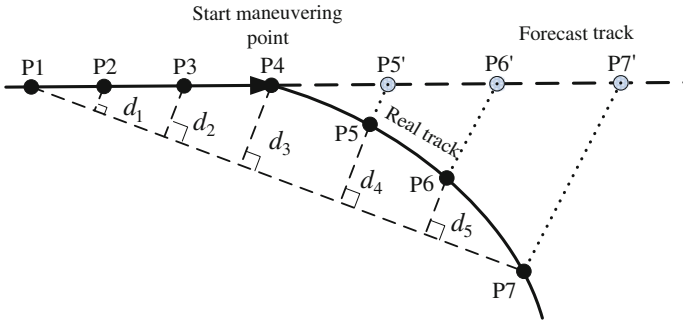


Fig. 93.2 The principle of algorithm

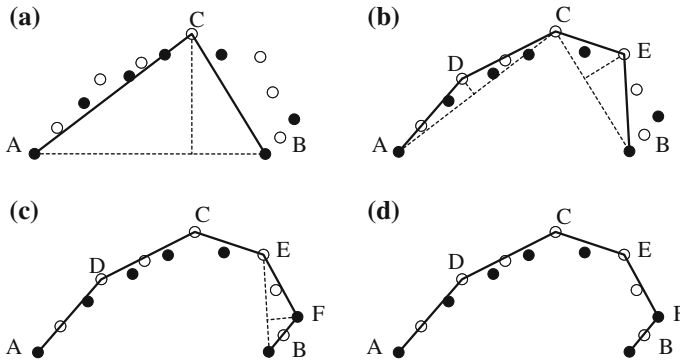
On the base of the target maneuver has been determined, the number of points which used to be processed by iterative endpoint fitting method would be choose according to certain principle, and then the target maneuvering information can be abstracted on accordance with the relationship that the distances of the points to the line which is made of start point and the end point.

### 93.3 Maneuver Target Detection Method with Iterative Endpoint Fitting Assisted

#### 93.3.1 Identification of Target Maneuvering

Set the measurement error of the sensor in the  $x$  direction is variance  $d_x$ , in the  $y$  direction is variance  $d_y$ ; the difference between observed value and predicted value in  $x$  direction is variance  $d'_x$ , in the  $y$  direction is variance  $d'_y$ . It was found by experiments that when choose  $3d_x$  and  $3d_y$  as threshold values the best detection and tracking performance could be gotten, the target occurred maneuver discriminate in accordance with Formula (93.1), if (1), target maneuvered in the  $x$  direction, else if (2), target maneuvered in the  $y$  direction, else if (3), target maneuvered in the  $x$  direction and  $y$  direction.

$$\begin{cases} d'_x \geq 3d_x \text{ and } d'_y \leq 3d_y & (1) \\ d'_x < 3d_x \text{ and } d'_y \geq 3d_y & (2) \\ d'_x \geq 3d_x \text{ and } d'_y \geq 3d_y & (3) \end{cases} \quad (93.1)$$



**Fig. 93.3** The principle of iterative endpoint fitting method. **a** First step, **b** Second step, **c** Third step, **d** Final step

### 93.3.2 Iterative Endpoint Fitting Method

The iterative endpoint fitting method is a method of curve fitting of the discrete points in digital image processing, the principle is shown in Fig. 93.3, and the specific steps are as follows [8].

① Connect two endpoints of the set of discrete points  $A$  and  $B$  to get a straight line  $AB$ , calculate the distances of the points to  $AB$ ;

② Set a threshold value, find the point of maximum distance to the straight line  $AB$  as  $C$ , if the distance is less than the threshold, then the set of discrete points can be fitted by one same function, else go to step ③;

③ Connect point  $A$  with  $C$  and  $B$  with  $C$ , and then do iterative process follow the steps ① and ②, until find out all points that the distances to the line are greater than the threshold, which are called key points, and points between the two adjacent key points fitted by same function to obtain a final fitting function.

### 93.3.3 Target Maneuvering Detection

① Calculate the number of points be used to fit. If there are a group of points which all the differences between observed values and predicted values are greater than the measurement error of the sensor, and the differences of the point latest loaded is greater than threshold, set the number of these points as  $i$ , then the number used of points be used to be fitted which set as  $n$  could be calculate follow as Formula (93.2).

$$n = 2 \times i + 1 \quad (93.2)$$

② Iterative endpoint fitting process. Set the coordinates of the two endpoints as  $(x_s, y_s)$  and  $(x_e, y_e)$ , the coordinates of the discrete point as  $(x_i, y_i)$ , the formula of the straight could be calculated follow Formula (93.3) and the distances of the discrete points to the line can be calculated follow Formula (93.4).

$$\begin{cases} y = \frac{y_e - y_s}{x_e - x_s}x + y_s - \frac{y_e - y_s}{x_e - x_s}x_s & y_e \neq y \text{ and } x_e \neq x_s \\ x = x_0 & y_e \neq y \text{ and } x_e = x_s = x_0 \\ y = y_0 & y_e = y = y_0 \text{ and } x_e \neq x_s \end{cases} \quad (93.3)$$

$$d = \begin{cases} \frac{|\frac{y_e - y_s}{x_e - x_s}x_i - y_i + y_s - \frac{y_e - y_s}{x_e - x_s}x_s|}{\sqrt{(\frac{y_e - y_s}{x_e - x_s})^2 + 1}} & y_e \neq y \text{ and } x_e \neq x_s \\ |x_0 - x_i| & y_e \neq y \text{ and } x_e = x_s = x_0 \\ |y_0 - y_i| & y_e = y = y_0 \text{ and } x_e \neq x_s \end{cases} \quad (93.4)$$

③ Compare the distance of each point to the line to identify the point of maximum distance value, then the point can be determined as start maneuver point and switch the tracking model, proceed to Step ④.

④ Output the maneuver detection results of the current target, to determine whether all targets discriminate completed, if completed, the end, otherwise return to Step ① to process next target.

Through an iterative endpoint fit method to get target maneuver information, and feedback the information to the target maneuvering detection processing, the performance of target maneuvering detection could be improved.

### 93.3.4 Data Processing Flow

Data processing flow is shown in Fig. 93.4.

## 93.4 Simulation and Analysis of Algorithm

Set target motion plot as that a moving target in the horizontal plane is observed by a two coordinates radar, the target do linear motion along negative direction at the time of at the speed of, the start position of the target is, and at the time of to make a turn to positive direction, the acceleration is, form the time of make a turn to negative direction, the acceleration is until to the time of. Radar scan period is and in the direction and direction observe independently, observation noise standard deviation were all.

In order to verify the effectiveness of the algorithm, in response the above scene, the comparative calculation tracking algorithm based on singer model and tracking algorithm based on singer model with iterative endpoint fitting assisted has been done, target tracking tracks are shown in Fig. 93.5.

Figure 93.6 shows comparisons of the estimated square errors in the direction and direction of two methods.

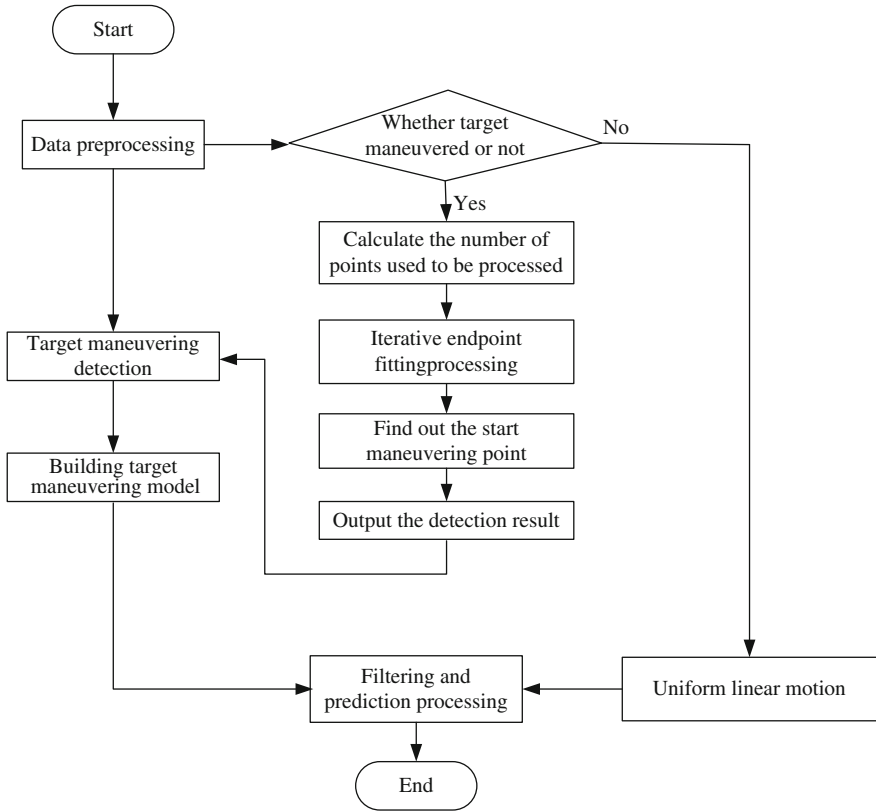
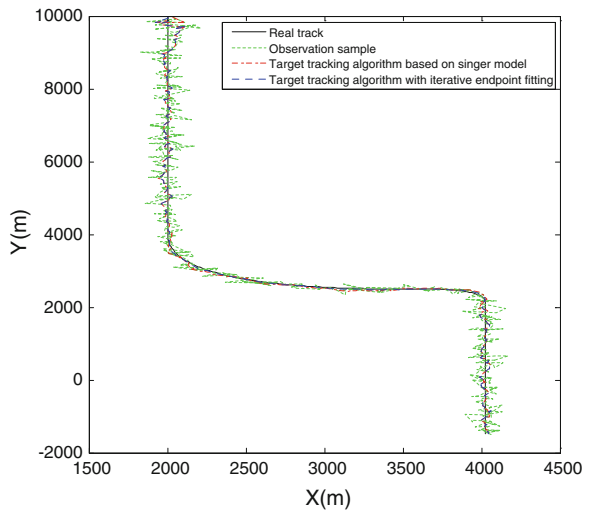


Fig. 93.4 Data processing flows

Fig. 93.5 Target tracking tracks of the two methods



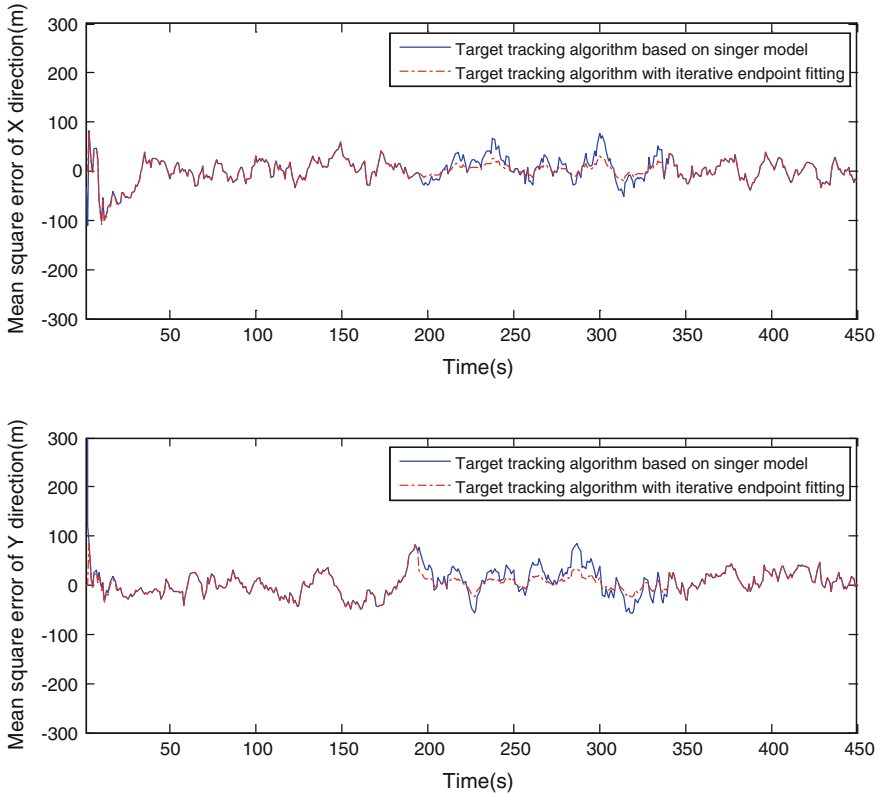


Fig. 93.6 Comparisons of the estimated square errors of two methods

As can be seen from the simulation results, with the detection method assisted which is proposed in this paper, when Singer model is used for target tracking, the performances of detection and tracking are better, and filtering accuracy is more accurate, the accuracy and practical of the algorithm has been verified in theory.

### 93.5 Conclusions

Maneuvering target tracking problem has been one of the difficult problems in the field of target tracking. How to improve the fast performance and tracking performance of the tracking method is the key to maneuvering target tracking algorithm research. In this paper, we consider the similarities of the curve fitting with the target maneuvering track, drawing on the iterative endpoint fit method in image processing, put forward a target maneuvering auxiliary detection method to try to improve the performance of maneuvering target tracking, which expanded



target maneuver detection method research ideas, has certain theoretical significance and application reference value.

## References

1. Fan HQ, Wang S, Fu Q (2009) Survey of algorithms of target maneuver detection. *Syst Eng Electron* 31(5):1064–1069 (In Chinese)
2. Fan HQ, Zhu YL, Fu Q (2011) Impact of mode decision delay on estimation error for maneuvering target interception. *IEEE Trans Aerosp Electron Syst* 47(1):702–711
3. Bizup DF, Brown DB (2004) Maneuver detection using the range rate measurement. *IEEE Trans AES* 40(1):330–336
4. Ru JF, Chen HM, Li XR et al (2005) A range based detection technique for tracking a maneuvering target. In: *Proceedings of SPIE on signal and data processing of small targets*
5. Chan YT, Hu AGC, Plant JB (1979) A Kalman filter based tracking scheme with input estimation. *IEEE Trans AES* 15(2):237–244
6. Li XR, Vesselin PJ (2002) A survey of maneuvering target tracking-Part IV: decision-based methods. In: *Proceedings of SPIE conference on signal and data processing of small targets*
7. Shi ZS, Xie J (2007) Simulation and analysis of the characters of maneuvering detection algorithm. *Comput Simul* 24(9):90–94 (In Chinese)
8. Yu SY, Zhou YH, Zhang R (2007) *Digital image processing*. Shanghai Jiao tong University Press, Shanghai (In Chinese)

# Chapter 94

## Multi-Camera Tracking via Online Discriminative Feature and Multi-Cue MRF

Jianyong Wang, Feng Chen, Jianwu Dong and Dingcheng Feng

**Abstract** Visual tracking across distributed cameras with disjoint views consists of many challenges, such as illumination changing and similar appearance of multiple persons. In this paper, we present a new solution to the problem in the formulation of Multi-Cue Markov Random Field (Multi-Cue MRF), and employ the max-product linear programming (MPLP) algorithm to find the MAP configuration of MRF. Moreover, in order to bridge the gap among different camera views, we propose a hybrid strategy which integrates spatio-temporal relationship modeling, online visual feature selection and local pair-wise code (LPWC) extraction into one framework. Finally, experimental results conducted with challenging video sequences verify the effectiveness of our method.

**Keywords** Multi-camera tracking · Multi-cue MRF · Surveillance

### 94.1 Introduction

Multi-camera tracking is significant in many areas such as building safety and security. It is a very challenging problem despite of large amount of study on it. In practical applications, the distributed cameras with non-overlapping views are usually required due to physical constraints, which leads to different view angles and lighting conditions from different cameras. Javed et al. [1] used a kernel density estimation approach to model the spatio-temporal relationship, and trained brightness transfer functions (BTFs) to handle the color transformation. However,

---

J. Wang · F. Chen (✉) · J. Dong · D. Feng  
Department of Automation, Tsinghua University, Beijing 100084, China  
e-mail: chenfung@mail.tsinghua.edu.cn

J. Wang  
e-mail: jy-wang06@mails.tsinghua.edu.cn

the illumination usually changes along with time, so the appearance of the same object can look much different due to the non-uniform illumination conditions, automatic white balance or infrared radiation, which the BTFs cannot explicitly cover. Matei et al. [2] combined joint kinematic and appearance features to solve a multi-hypothesis tracking problem. Song et al. [3] viewed similarity between neighboring views as a random variable and estimated its distribution. They also measured the feature variance along the estimated path to avoid association error. Kuo et al. [4] proposed an appearance affinity model by online multiple instance learning. However, the trained classifier lacks of generalization ability for new instances. Practically, we only need to solve the association within a sliding time window, and hence it will be beneficial to select the most discriminative features to separate the objects from different periods and camera views. The similar idea can be found in [5], where the authors projected features into a subspace by partial least squares reduction using a one-against-all scheme. Picus et al. [6] employed a branch-and-bound technique to construct trajectories association based on geometric relations with calibrated cameras. Mazzon et al. [7] assumed that the people's behavior towards different goals can be predicted by a motion model without any appearance features. To coordinate labels for two cameras efficiently, a natural way is to formulate it as a weighted bipartite graph, which can be solved with complexity  $O(n^{2.5})$  [1], but the problem is NP hard for more than two cameras. To handle this case, Jiang et al. [8] formulated the problem as a multi-path covering problem and solved it by LP relaxation. Chen et al. [9] also proposed a Markov Chain Monte Carlo (MCMC) method.

In this paper, we address the task by combining spatio-temporal and online discriminative features in the Multi-Cue MRF framework, which is a popular and powerful tool in modeling relationship among large variable and is widely used in the field of computer vision. For spatio-temporal features, we assume that the transition distance in each blind area can be modeled by independent Gaussian distributions. The appearance feature includes two novel parts: (i) we identify the global optimal feature subset adaptively for different people sets by a trace ratio criterion [10], which is crucial when people have similar dressing or the lighting condition is poor; (ii) we introduce local pair-wise code (LPWC) for pairs in the same view to penalize the error for hypothesized associations, as we assume that the relative appearance relations are more robust to illumination.

Our main contributions include: (i) we perform online feature selection to preserve the most powerful appearance feature subset; (ii) we design the LPWC to suppress the illumination-originated bad impact; (iii) we use Multi-Cue MRF to solve the problem with a well-known inference algorithm named MPLP [11], making the computation process more systematically and efficiently.

## 94.2 Formulation Using Multi-Cue MRF

We aim to coordinate persons passing through multiple cameras with non-overlapping views based on the detection and tracking results in single cameras. Suppose the camera network topology is available, and there are  $N$  complete passing through (observations) in the entrance regions with unique labels  $L_0 = \{l_1, l_2, \dots, l_N\}$ , and we denote labels of observations from other cameras by  $L = \{l'_1, l'_2, \dots, l'_M\}$ . Our task is to assign labels for  $\mathbf{L}$  from the candidate set  $\mathbf{L}_0$ . Note that each observation is comprised of  $Q$  snapshots with original image pixels and transition time/position/velocity information when appearing/disappearing. In the following we denote  $i^{\text{th}}$  observation with unknown label as  $O_i$ .

As the number of associations increases rapidly when person or camera number becomes larger, it is difficult to achieve the global optimum. In our work, we approximate the solution by MRF, which is a class of well-known graphical models. In MRF, many existing inference algorithms can be used to solve the MAP problem efficiently. Moreover, the MRF modeling is flexible with different camera setup contexts. Suppose there is a hidden variable vector  $x$  and observation vector  $y$ , the distribution  $P(x|y)$  can be factorized to the product of node potentials and edge potentials defined on cliques of the graph

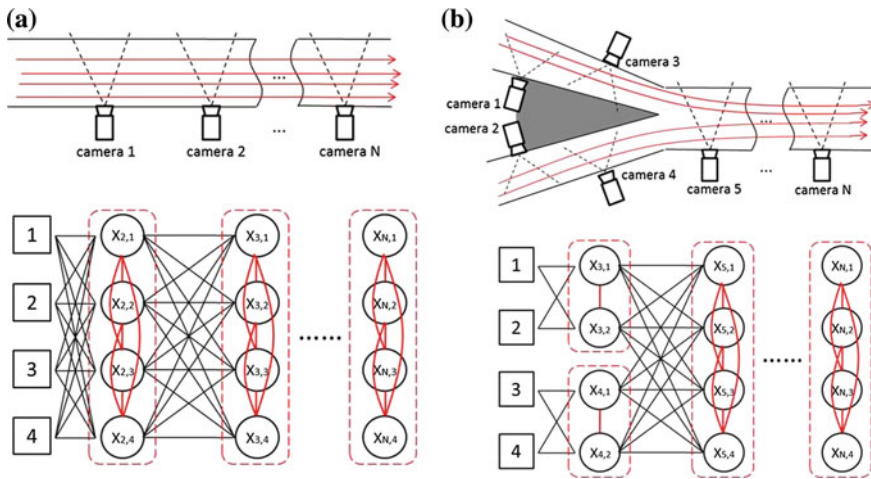
$$p(x|y) \propto \prod_{k \in \text{Vertex}} \psi(x_k) \prod_{(i,j) \in \text{Edge}} \psi(x_i, x_j). \quad (94.1)$$

The objective can be

$$\begin{aligned} x^* &= \arg \max_x \prod_{k \in \text{Vertex}} \psi(x_k) \prod_{(i,j) \in \text{Edge}} \psi(x_i, x_j) \\ &= \arg \min_x \sum_{k \in \text{Vertex}} \theta(x_k) + \sum_{(i,j) \in \text{Edge}} \theta(x_i, x_j), \end{aligned} \quad (94.2)$$

where  $\theta(x_k)$  and  $\theta(x_i, x_j)$  is the negative logarithm of node and edge potential. For simplicity, we name them by node potential and edge potential, respectively.

When a camera encounters an observation, a vertex will be added into the graphical model. The node potential can be interpreted as the penalty matrix when the observation is assigned by different labels, and the edge potential is the penalty matrix when the two linked vertices are assigned by different joint labels. For  $O_i$  its previous and successive observations determined by the topology and transition time threshold are represented by  $pre(O_i)$  and  $suc(O_i)$ . In the Multi-Cue MRF, the successive observations of an observation form a block where the vertices are fully connected. The edges in the block enforce the mutually exclusive labels and evaluate assignment by LPWC, respectively. Other edges connect the possible transition pairs in adjacent cameras. Examples of Multi-Cue MRF are in Fig. 94.1.



**Fig. 94.1** Multi-Cue MRF for different cases, the dotted box indicates the vertex block. **a** 4 persons pass through cameras in a chain. **b** 4 persons pass through cameras from two entrance regions

### 94.3 Spatio-Temporal and Appearance Feature Relationship Modeling

The transition time interval and velocity are commonly used cues for association. Another important cue is the color feature, and its similarity comparison can be implemented in a more discriminative and lower dimensional space in our method. Additionally, LPWC is proposed based on the fact that the relative appearance will be maintained for pairs in the same view.

#### 94.3.1 Spatio-Temporal Feature

Let the spatio-temporal state of  $O_i$  be  $St(O_i) = [V^a(O_i), V^d(O_i), T^a(O_i), T^d(O_i)]$ , which includes the appearing/disappearing velocity and time. The Gaussian distribution  $N(\bullet)$  of transition distance can be learnt by the state vector of the same person. Thus, the transition penalty measured by spatio-temporal feature is

$$f^{st}(O_i, O_j) = 1 - N_{i,j}(0.5(V^a(O_j) + V^d(O_i)) \cdot (T^a(O_j) - T^d(O_i))). \quad (94.3)$$

### 94.3.2 Discriminative Appearance Feature Selection

Although the spatio-temporal relationship is effective if people traverse in order, the appearance features could be more discriminative when more persons interact with the system together. We select the channels in HSV and YCrCb space (V and Y are the same) to calculate the pixel average of horizontal stripe. All the image channels are normalized to the size of  $60 \times 100$ . Thereby, one snapshot's feature is represented by a normalized 500 dimensional vector. Suppose we seek the subset from P tracks with Q snapshots, then the feature matrix is  $X = [x_1, x_1, \dots, x_{PQ}]$ , where each column is the feature vector. We denote the optimal subset by  $Y = [y_1, y_1, \dots, y_{PQ}]$ , and selection matrix by  $W$ , then we have  $Y = W^T X$ . In [9], the within-class and between-class matrix  $A_w$  and  $A_b$  in Fisher score scheme are

$$(A_w)_{ij} = \begin{cases} \frac{1}{Q} & l(x_i) = l(x_j) \\ 0 & \text{else} \end{cases} \quad (A_b)_{ij} = \begin{cases} \frac{1}{PQ} - \frac{1}{Q} & l(x_i) = l(x_j) \\ \frac{1}{PQ} & \text{else} \end{cases}, \quad (94.4)$$

where  $l(x_i)$  represents the class of  $x_i$ . The selection criteria can be written as

$$W^* = \arg \max_w \frac{\sum_{ij} \|y_i - y_j\|^2 (A_b)_{ij}}{\sum_{ij} \|y_i - y_j\|^2 (A_w)_{ij}} = \arg \max_w \frac{\text{trace}(W^T X (D_b - A_b) X^T W)}{\text{trace}(W^T X (D_w - A_w) X^T W)}, \quad (94.5)$$

where  $D_b$  and  $D_w$  are diagonal matrices, and  $(D_b)_{ii} = \sum_j (A_b)_{ij}$ ,  $(D_w)_{ii} = \sum_j (A_w)_{ij}$ . An algorithm [9] is used to find the global optimal subset. Finally, if the subset is selected based on observation set  $\mathbf{O}$ , the similarity can be measured by Bhattacharyya distance

$$f^{app}(O_i, O_j, \mathbf{O}) = \text{Bhattacharyya}(Y^o(O_i), Y^o(O_j)), \quad (94.6)$$

where  $Y^o(O_i)$  is the average feature of  $O_i$ 's snapshots.

### 94.3.3 Local Pair-Wise Code Extraction

In most surveillance tasks, the illumination changing arises from lots of reasons and it is not feasible to model them all explicitly. In this section, the LPWC is defined to coarsely depict the invariant relationship, which is assumed that two persons will reserve relative appearance as long as they are in the same view. For example, person A's jacket is darker than person B's, while pants are lighter in camera M, this relationship will be kept in camera N. To extract LPWC, the body is firstly divided into three regions through a fixed ratio: the head, torso and leg,

and then in each part the average pixel for all the snapshots will be calculated. For  $O_i$ , the concatenated average values in all parts and channels is denoted by  $Z_i$ . The LPWC calculation flowchart is explained in Algorithm 1 and Fig. 94.2.

---

**Algorithm 1** Calculate  $lpwc(O_i, O_j)$

---

```

S = null
if  $O_i, O_j$  belong to same camera view then
  for each dim d do
    if  $(Z_i[d] - Z_j[d]) / (\min(Z_i[d], Z_j[d]) + \epsilon) > \tau$  then
      S[d] = 1
    else if  $(Z_i[d] - Z_j[d]) / (\min(Z_i[d], Z_j[d]) + \epsilon) < -\tau$  then
      S[d] = -1
    else
      S[d] = 0
    end if
  end for
end if
 $lpwc(O_i, O_j) = S$ 

```

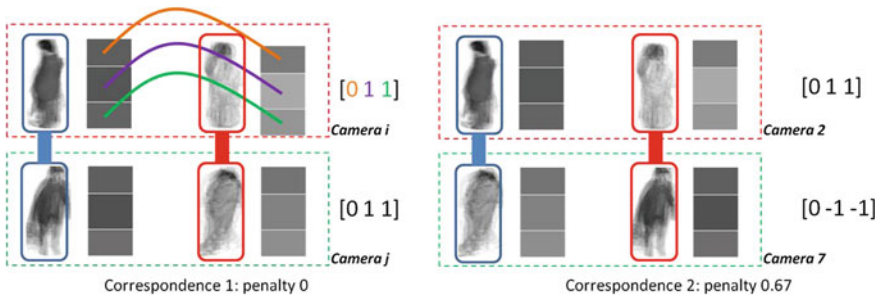
---

The penalty when associating  $O_i$  with  $O_m$  and  $O_j$  with  $O_n$  is measured by Hamming distance (if one of the input  $lpwc$  is null, the result is 0).

$$f^{lpwc}(O_m, O_n, O_i, O_j) = Hamming(lpwc(O_m, O_n), lpwc(O_i, O_j)). \tag{94.7}$$

### 94.3.4 Potential Functions

In this subsection, we focus on how to choose the potential functions of Multi-Cue MRF. Let the initial observation set with known labels be  $\mathbf{V}$ , then the  $i^{th}$  element in the node potential is the penalty when this vertex's state is assigned by the label of  $V_i$ . Let the  $k^{th}$  vertex be  $V_k$ . For simplicity, we assume the vertex is the same with observation. Therefore, the node potential can be written as



**Fig. 94.2** The process of LPWC calculation. The  $lpwc$  is only calculated in gray scale for illustration

$$(\theta(x_k))[i] = \alpha f^{app}(\mathbf{V}_i, V_k, \mathbf{V}) + \beta f^{st}(\mathbf{V}_i, V_k) \cdot \delta(V_k \in pre(\mathbf{V}_i)). \quad (94.8)$$

The edge potential indicates the joint penalty when the vertex pair is assigned by different joint states, and it can be divided into two types in Multi-Cue MRF: the inner edge potential and the outer edge potential. The inner edge potential is defined inside the vertex blocks:

$$(\theta(x_i, x_j))[m, n] = \begin{cases} f^{lpwc}(\mathbf{V}_m, \mathbf{V}_n, V_i, V_j) & m \neq n \\ \varphi & m = n, \end{cases} \quad (94.9)$$

here  $\varphi$  is used to penalize the assignment of the same label to different observations in the block. The outer edge potential is defined on the edge between adjacent cameras to measure the transition penalty:

$$(\theta(x_i, x_j))[m, n] = \begin{cases} \lambda f^{st}(V_i, V_j) + \mu f^{app}(V_i, V_j, pre(V_j)) & m \neq n \\ \gamma & m = n, \end{cases} \quad (94.9)$$

where  $\gamma$  is to penalize wrong label assignment for branch paths. Finally, we solve the problem in cluster-based LP relaxation scheme. To achieve a good tradeoff between accuracy and computational complexity, MPLP [10] is chosen to solve the cluster adding problem with the well-known dual LP method.

## 94.4 Experimental Results

We conducted experiments in two typical video sequences: *elevator* and *building*. In *elevator*, there are eight cameras on eight floor lobbies, and one zenithal camera inside the elevator car. The lighting condition has dramatic change because lamp number and window position differ among lobbies, which renders targets almost indistinguishable even by human eyes. In the busiest time, there are eight passengers transferring together. In *building*, there are five cameras mounted outdoors, indoors and on different floors. To test our method in more challenging situations, we selected the videos taken in winter, when the dark and heavy clothes make people appearance look very similar. Note that the appearance similarity can be ignored if view angles are totally different, such as the views in the elevator lobby and car. The experimental results are shown in Fig. 94.3 and Table 94.1.

From Fig. 94.4 we can see that the accuracy decreases when the LPWC or feature selection strategy is dropped. In *elevator*, the illumination changing is fiercely due to infrared radiation, lamp number, etc., but the clothes are a bit more diverse, so dropping LPWC will hurt accuracy more. Unlike *elevator*, persons in *building* are dark, leading to less discriminative relative appearance. However, the illumination did not vary that fiercely, so the selected local features play more important roles.



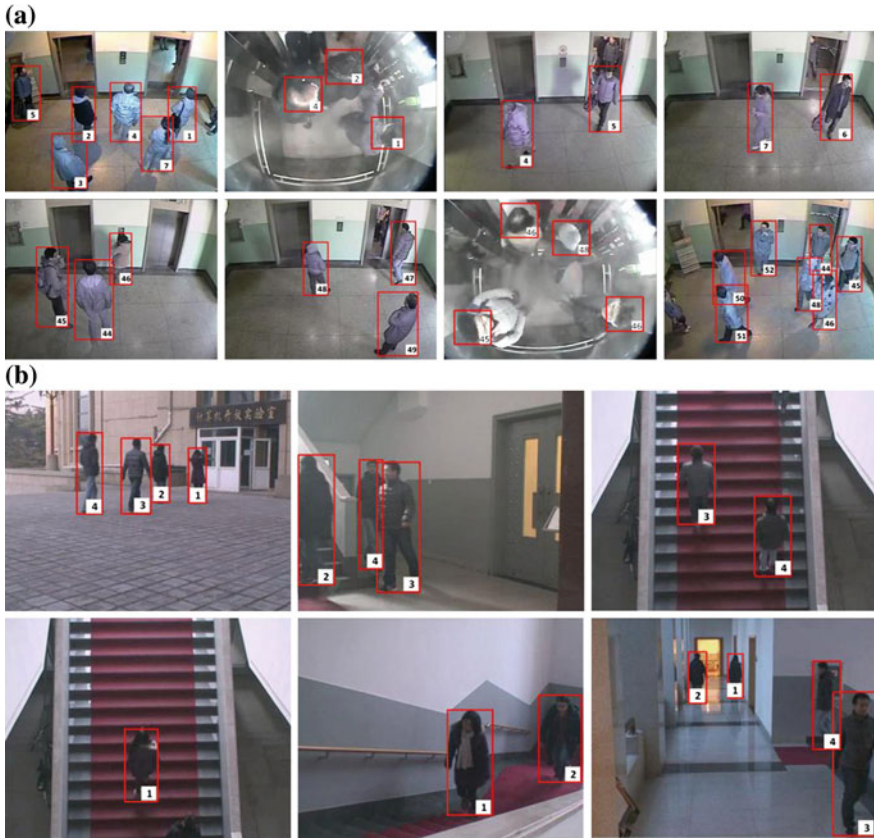


Fig. 94.3 Experimental results: **a** elevator video sequence; **b** building video sequence

**Table 94.1** Multi-camera correspondence result in two video sequences

Video seq.	# of camera	# of transition	# of correct correspondence
<i>Elevator</i>	8	98	94
<i>Building</i>	5	40	40

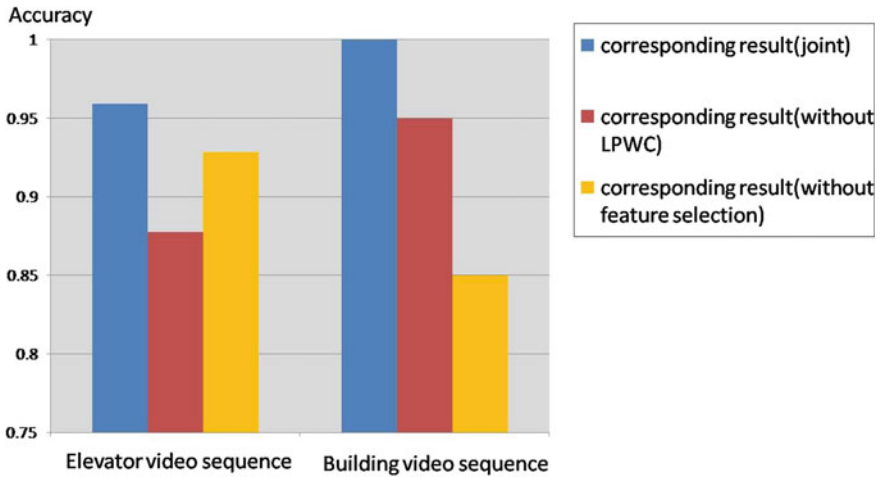


Fig. 94.4 Accuracy comparison of our method to the results without LPWC or feature selection

### 94.5 Conclusion

In this paper, we propose a method to solve the person correspondence problem among multi-cameras, where a Multi-Cue MRF framework with MPLP algorithm is employed. In order to make the solution more robust and adaptive, we utilize a hybrid method which combines spatio-temporal relationship modeling, online discriminative feature selection strategy and local pair-wise appearance code extraction. Our future work includes incorporating it into intra-camera detection and tracking systems.

**Acknowledgments** This work was supported by National Natural Science Foundation of China (Project No.61071131 and Project No.61271388), Beijing Natural Science Foundation (No.4122040), Research Project of Tsinghua University (No.2012Z01011) and Doctoral Fund of Ministry of Education of China (No. 20120002110036).

### References

1. Javed O, Shafique K, Rasheed Z, Shah M (2008) Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views. *CVIU* 109(2):146–162
2. Matei B, Sawhney H, Samarasekera S (2011) Vehicle tracking across nonoverlapping cameras using joint kinematic and appearance features. In: *CVPR*, pp 3465–3472
3. Song B, Roy-Chowdhury A (2008) Robust tracking in a camera network: a multi-objective optimization framework. *JSTSP* 2(4):582–596
4. Kuo C, Huang C, Nevatia R (2010) Inter-camera association of multi-target tracks by on-line learned appearance affinity models. In: *ECCV*, pp 383–396
5. Schwartz W, Davis L (2009) Learning discriminative appearance-based models using partial least squares. In: *SIBGRAPI*, pp 322–329

6. Picus C, Pflugfelder R, Micusik B (2011) Branch and bound global optima search for tracking a single object in a network of non-overlapping cameras. In: ICCV, pp 1825–1830
7. Mazzon R, Cavallaro A (2012) Multi-camera tracking using a multi-goal social force model. *Neurocomputing* 5(2):1–12
8. Jiang H, Fels S, Little J (2007) A linear programming approach for multiple object tracking. In: CVPR, pp 1–8
9. Chen K, Lai C, Hung Y, Chen C (2008) An adaptive learning method for target tracking across multiple cameras. In: CVPR, pp 1–8
10. Nie F, Xiang S, Jia Y, Zhang C, Yan S (2008) Trace ratio criterion for feature selection. *AAAI* 2:671–676
11. Sontag D, Meltzer T, Globerson A, Jaakkola T, Weiss Y (2008) Tightening lp relaxations for map using message passing. In: UAI