



Computer Science and Information Systems

Published by ComSIS Consortium

**Special Issue on Advances in Computer
Animation and Digital Entertainment**

Volume 7, Number 1
February 2010

ComSIS Consortium:

University of Belgrade:

Faculty of Organizational Science, Belgrade, Serbia

Faculty of Mathematics, Belgrade, Serbia

School of Electrical Engineering, Belgrade, Serbia

Serbian Academy of Science and Art:

Mathematical Institute, Belgrade, Serbia

Union University:

School of Computing, Belgrade, Serbia

University of Novi Sad:

Faculty of Sciences, Novi Sad, Serbia

Faculty of Technical Sciences, Novi Sad, Serbia

Faculty of Economics, Subotica, Serbia

Technical Faculty "M. Pupin", Zrenjanin, Serbia

University of Montenegro:

Faculty of Economics, Podgorica, Montenegro

Faculty of Mathematics and Science, Podgorica, Montenegro

EDITORIAL BOARD:

Editor-in-Chief: Mirjana Ivanović, University of Novi Sad

Vice Editor-in-Chief: Ivan Luković, University of Novi Sad

Managing Editor: Zoran Putnik, University of Novi Sad

Managing Editor: Miloš Radovanović, University of Novi Sad

Editorial Assistant: Vladimir Kurbalija, University of Novi Sad

Editorial Assistant: Jovana Vidaković, University of Novi Sad

Editorial Assistant: Ivan Pribela, University of Novi Sad

Editorial Assistant: Slavica Aleksić, University of Novi Sad

Editorial Assistant: Srđan Škrbić, University of Novi Sad

Associate Editors:

P. Andreae, Victoria University, New Zealand

D. Bojić, University of Belgrade, Serbia

D. Bečejski-Vujaklija, University of Belgrade, Serbia

I. Berković, University of Novi Sad, Serbia

G. Bhutkar, Vishwakarma Institute of Technology, India

L. Böszörményi, University of Clagenfurt, Austria

K. Bothe, Humboldt University of Berlin, Germany

Z. Budimac, University of Novi Sad, Serbia

G. Devedžić, University of Kragujevac, Serbia

V. Devedžić, University of Belgrade, Serbia

J. Đurković, University of Novi Sad, Serbia

S. Guttormsen Schar, ETH Zentrum, Switzerland

P. Hansen, University of Montreal, Canada

L. C. Jain, University of South Australia, Australia

V. Jovanović, Georgia Southern University, USA

Lj. Kaščelan, University of Montenegro, Montenegro

P. Mahanti, University of New Brunswick, Canada

M. Maleković, University of Zagreb, Croatia

N. Mitić, University of Belgrade, Serbia

A. Mitrović, University of Canterbury, New Zealand

N. Mladenović, Serbian Academy of Science, Serbia

P. Mogin, Victoria University, New Zealand

S. Mrdalj, Eastern Michigan University, USA

G. Nenadić, University of Manchester, UK

Z. Ognjanović, Serbian Academy of Science, Serbia

A. Pakstas, London Metropolitan University, England

P. Pardalos, University of Florida, USA

M. Racković, University of Novi Sad, Serbia

B. Radulović, University of Novi Sad, Serbia

M. Stanković, University of Niš, Serbia

D. Tošić, University of Belgrade, Serbia

J. Trninić, University of Novi Sad, Serbia

J. Woodcock, University of Kent, England

EDITORIAL COUNCIL:

S. Ambroszkiewicz, Polish Academy of Science, Poland

Z. Arsovski, University of Kragujevac, Serbia

D. Banković, University of Kragujevac, Serbia

T. Bell, University of Canterbury, New Zealand

S. Bošnjak, University of Novi Sad, Serbia

Ž. Branović, University of Novi Sad, Serbia

H. D. Burkhard, Humboldt University of Berlin, Germany

B. Chandrasekaran, Ohio State University, USA

V. Ćirić, University of Belgrade, Serbia

D. Domazet, Faculty of Information Technology, Belgrade, Serbia

S. Đrođević-Kajan, University of Niš, Serbia

P. Hotomski, University of Novi Sad, Serbia

P. Janičić, University of Belgrade, Serbia

Z. Jovanović, University of Belgrade, Serbia

L. Kalinichenko, Russian Academy of Science, Russia

Z. Konjović, University of Novi Sad, Serbia

I. Koskosas, University of Western Macedonia, Greece

W. Lamersdorf, University of Hamburg, Germany

T. C. Lethbridge, University of Ottawa, Canada

A. Lojpur, University of Montenegro, Montenegro

Y. Manolopoulos, Aristotle University, Greece

A. Mishra, Atilim University, Turkey

S. Mishra, Atilim University, Turkey

J. Protić, University of Belgrade, Serbia

D. Simpson, University of Brighton, England

D. Starčević, University of Belgrade, Serbia

S. Stojanov, Faculty for Informatics, Plovdiv, Bulgaria

D. Surla, University of Novi Sad, Serbia

M. Tuba, University of Belgrade, Serbia

P. Tumbas, University of Novi Sad, Serbia

P. Zarate, IRIT-INPT, Toulouse, France

K. Zdravkova, University of Ss. Cyril and Methodius, Skopje, FYROM

ComSIS Editorial office:

**University of Novi Sad, Faculty of Sciences,
Department of Mathematics and Informatics**
Trg Dositeja Obradovica 4, 21000 Novi Sad, Serbia
Phone: +381 21 458 888; **Fax:** +381 21 6350 458
www.comsis.org; Email: comsis@uns.ac.rs

**Volume 7, Number 1, Special Issue, 2010
Novi Sad**

Computer Science and Information Systems

ISSN: 1820-0214

ComSIS Journal is sponsored by:

Ministry of Science and Technological Development of Republic of Serbia -
<http://www.nauka.gov.rs/lat>



Computer Science and Information Systems

AIMS AND SCOPE

Computer Science and Information Systems (ComSIS) is an international refereed journal, published in Serbia. The objective of ComSIS is to communicate important research and development results in the areas of computer science, software engineering, and information systems.

We publish original papers of lasting value covering both theoretical foundations of computer science and commercial, industrial, or educational aspects that provide new insights into design and implementation of software and information systems. ComSIS also welcomes surveys papers that contribute to the understanding of emerging and important fields of computer science. Regular columns of the journal cover reviews of newly published books, presentations of selected PhD and master theses, as well as information on forthcoming professional meetings. In addition to wide-scope regular issues, ComSIS also includes special issues covering specific topics in all areas of computer science and information systems.

ComSIS publishes invited and regular papers in English. Papers that pass a strict reviewing procedure are accepted for publishing. The acceptance rate so far was approximately 40%. ComSIS is published semiannually.

Indexing Information

ComSIS is indexed and abstracted in the following:

- Science Citation Index Expanded (also known as SciSearch) and Journal Citation Reports / Science Edition by Thomson Reuters,
- Computer Science Bibliography, University of Trier (DBLP),
- EMBASE (Elsevier),
- Scopus (Elsevier),
- Summon (Serials Solutions),
- Selected for coverage in EBSCO bibliographic databases,
- Selected for coverage in IET bibliographic database Inspec, starting with 2009 published material,
- Selected for coverage in FIZ Karlsruhe bibliographic database io-port,
- Google Scholar,
- Center for Evaluation in Education and Science and Ministry of Science of Republic of Serbia (CEON) in cooperation with the National Library of Serbia,
- Serbian Citation Index (SCIndeks),
- doiSerbia.

Information for Contributors

The Editors will be pleased to receive contributions from all parts of the world. An electronic version (MS Word), or three hard-copies of the manuscript written in English, intended for publication and prepared as described in "Manuscript Requirements" (which may be downloaded from <http://www.comsis.org>), along with a cover letter containing the corresponding author's details should be sent to one of the Editors.

Criteria for Acceptance

Criteria for acceptance will be appropriateness to the field of Journal, as described in the Aims and Scope, taking into account the merit of the content and presentation. There is no page limit on manuscripts submitted.

Manuscripts will be refereed in the manner customary with scientific journals before being accepted for publication.

Copyright and Use Agreement

All authors are requested to sign the "Transfer of Copyright" agreement before the paper may be published. The copyright transfer covers the exclusive rights to reproduce and distribute the paper, including reprints, photographic reproductions, microform, electronic form, or any other reproductions of similar nature and translations. Authors are responsible for obtaining from the copyright holder permission to reproduce the paper or any part of it, for which copyright exists.

Computer Science and Information Systems

Volume 7, Number 1, Special Issue, February 2010

CONTENTS

Editorial

Guest Editorial

Papers

- 1** **An Extended Photometric Stereo Algorithm for Recovering Specular Object Shape and Its Reflectance Properties**
Zuoyong Zheng, Lizhuang Ma, Zhong Li, and Zhihua Chen
- 13** **Extracting PCB Components Based on Color Distribution of Highlight Areas**
Zhou Zeng, LiZhuang Ma, and ZuoYong Zheng
- 31** **Service-oriented Enterprise Interoperability in Automobile Supply Chain Management**
Yong Zhang, Shijun Liu, Lei Wu, and Yuchang Jiao
- 51** **Modeling and Simulation of a Spherical Mobile Robot**
SANG Shengju, ZHAO Jichao, WU Hao, CHEN Shoujun, and AN Qi
- 63** **3D Mesh Skeleton Extraction Using Prominent Segmentation**
Xiaopeng SUN, J. PAN, and Xiaopeng WEI
- 75** **A Dynamic Alignment Algorithm for Imperfect Speech and Transcript**
Ye Tao, Xueqing Li, and Bian Wu
- 85** **Multi-Video Summarization Using Complex Graph Clustering and Mining**
Jian Shao, Dongming Jiang, Mengru Wang, Hong Chen, and Lu Yao
- 99** **A Retrieval Method for Human Mocap Data Based on Biomimetic Pattern Recognition**
Xiaopeng Wei, Boxiang Xiao, and Qiang Zhang
- 111** **Generative 3D Images in a Visual Evolutionary Computing System**
Hong Liu
- 127** **Effective Semi-supervised Nonlinear Dimensionality Reduction for Wood Defects Recognition**
Zhao Zhang and Ning Ye
- 139** **MFI-Tree: An Effective Multi-feature Index Structure for Weighted Query Application**
Yunfeng He and Junqing Yu

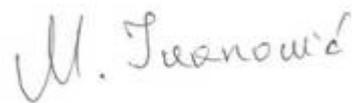
- 153 A Content-based Dynamic Load-Balancing Algorithm for Heterogeneous Web Server Cluster**
Zhang Lin, Li Xiao-ping, and Su Yuan
- 163 Blind Separation Using Second Order Statistics for Non-stationary Signals**
Jun Du, Ju Liu, Shengju Sang, and Jun Wang
- 177 An Accelerometer-Based Gesture Recognition Algorithm and its Application for 3D Interaction**
Jianfeng Liu, Zhigeng Pan, and Xiangcheng Li
- 189 The Design and Evaluation of Hierarchical Multi-level Parallelisms for H.264 Encoder on Multi-core Architecture**
Haitao Wei, Junqing Yu, and Jiang Li
- 201 Robust Moving Object Detection Under Complex Background**
DING Ying, LI Wen-hui, FAN Jing-tao, and YANG Hua-min
- 211 A Novel Hierarchical Speech Emotion Recognition Method Based on Improved DDAGSVM**
Qi-rong Mao and Yong-zhao Zhan
- 223 Research on Optimizing the Fault Diagnosis Strategy of Complex Electronic Equipment**
WANG Hong-xia, YE Xiao-hui, and WANG Liang
- 231 3D Point Pattern Matching Based on Spatial Geometric Flexibility**
Xiaopeng Wei, Xiaoyong Fang, Qiang Zhang, and Dongsheng Zhou

EDITORIAL

In the seventh year of publishing, we have the great pleasure to announce that this ComSIS has been selected for coverage in EBSCOhost's databases, adding another established vendor to the growing list of indexing databases where ComSIS is featured.

This is one of the special issues of ComSIS, devoted to advances in computer animation and digital entertainment. On behalf of the ComSIS Consortium, we would like to use this opportunity to thank the reviewers and all of the authors for their diligence and high-quality work. Most prominently, we would like to thank Qiang Zhang, the guest editor of this issue, as well as Chunlin Ma and Lizhuang Ma, general chairs of the 6th Conference on Intelligent CAD and Digital Entertainment (CIDE 2009), for their great contribution and efforts in organizing the conference and creating this issue of ComSIS.

Editor-in-Chief
Mirjana Ivanović



Managing Editor
Miloš Radovanović



GUEST EDITOR'S MESSAGE

With the rapid development of computer and network technologies in recent years, research and applications in computer-aided design (CAD), computer animation, digital entertainment and digital art have been receiving increasing attention from researchers across the world. The aim of this special issue of ComSIS is to further promote the development of those areas to benefit their applications in various fields.

This special issue is a collection of 19 papers selected from the 426 submissions to the 6th Conference on Intelligent CAD and Digital Entertainment (CIDE 2009) held in Taian, Shandong, China, from 15 to 17 August 2009. All the papers included in this special issue have gone through a rigorous peer-review and revision process for their originality and quality.

This issue of ComSIS takes the topic of advances in computer animation and digital entertainment as a center. Many of the papers covered mostly theoretical research and various application domains, such as intelligent CAD, computer graphics, computer animation, and motion capture animation.

In photometric stereo, the existence of specularities hampers recovery of the normal map. To deal with this common reflective phenomenon, in the paper "An extended photometric stereo algorithm for recovering specular object shape and its reflectance properties", Zuoyong Zheng, Lizhuang Ma, Zhong Li and Zhihua Chen introduce a novel representation for specular reflection with a set of specular basis functions with different roughness values. This representation is suitable for any intensively or weakly specular object, and is introduced into the photometric stereo algorithm to recover both the surface shape and its reflectance properties. The reconstructed shapes and re-rendered images validate the proposed algorithm.

The paper "Extracting PCB components based on color distribution of highlight areas" by Zhou Zeng, Lizhuang Ma and ZuoYong Zheng investigates methodologies for locating and identifying the components on a printed circuit board (PCB) used for surface mount device inspection. The proposed scheme consists of two stages: solder joint extraction and protective coating extraction. Solder joints are extracted by first detecting all the highlight areas, and then recognizing and removing the invalid highlight areas. The sequence of color distribution as a new clue has been applied to clustering solder joints. Each protective coating is extracted by the positions of the clustered solder joints.

How to enhance interoperability between stakeholders and improve efficiency of supply chain management is the key issue that needs to be addressed in automobile industry. Yong Zhang, Shijun Liu, Lei Wu, Yuchang Jiao, Xiangxu Meng propose a methodology that provides a guide on how to establish interoperability between enterprises through a federated approach in “Service-oriented enterprise interoperability in automobile supply chain management”. An interoperability service platform is designed and delivered in the form of software as a service (SaaS). This paper introduces the specifications of the service platform and proposes an interactive framework which is used to establish interoperability between the service platform and other on-premise applications.

Spherical mobile robot (SMR) has been studied analytically and experimentally in the paper “Modeling and simulation of a Spherical Mobile Robot” by Shengju Sang, Jichao Zhao, Hao Wu, Shoujun Chen and Qi An. A novel design with an internal propulsion mechanism and mathematical models of the robot’s dynamics and kinematics is introduced.

Based on feature point and core extraction by the multi-dimensional scaling (MDS) transformation, Xiaopeng Sun, J. Pan and Xiaopeng Wei present a novel skeleton extraction algorithm in their paper “3D Mesh Skeleton Extraction Using Prominent Segmentation”.

The paper “A Dynamic Alignment Algorithm for Imperfect Speech and Transcript” by Ye Tao, Xueqing Li and Bian Wu presents a novel alignment approach for imperfect speech and the corresponding transcription. It starts with multi-stage sentence boundary detection in audio, followed by a dynamic programming based search, to find the optimal alignment and detect the mismatches at sentence level.

Multi-video summarization is a great theoretical and technical challenge due to the wider diversity of topics in multi-video than single-video as well as the multi-modality nature of multi-video over multi-document. Dongming Jiang, Jian Shao and Lu Yao in their paper “Multi-video summarization using complex graph clustering and mining” propose an approach to analyze both visual and textual features across a set of videos and to create a so-called circular storyboard composed of topic-representative keyframes and keywords.

Xiaopeng Wei, Boxiang Xiao and Qiang Zhang present a retrieval method for human Mocap (motion capture) data based on biomimetic pattern recognition in their paper “A retrieval method for human Mocap data based on biomimetic pattern recognition”. BVH rotation channels are extracted as features of motion for both the retrieval instance and the motion data. Several hyper sausage neurons are constructed according to the retrieval instance, and the

trained domain covered by these hyper sausage neurons can be considered as the distribution range of a same kind of motions.

The paper “Generative 3D Images in a visual evolutionary computing system” by Hong Liu presents a novel computer-aided design system which uses a computational approach to producing 3D images for stimulating the creativity of designers.

Dimensionality reduction is an important preprocessing step in high-dimensional data analysis without losing much intrinsic information. Zhao Zhang and Ning Ye in their paper “Effective semi-supervised nonlinear dimensionality reduction for wood defects recognition” consider the problem of kernel based semi-supervised nonlinear dimensionality reduction, which is called KNDR. It can project the data onto a set of ‘useful’ features and preserve the structure of labeled and unlabeled data as well as the constraints defined in the embedded spaces, under which the projections of the data can be effectively partitioned from each other.

Multi-feature index tree (MFI-Tree), a new indexing structure, is proposed to index multiple high-dimensional features of video data for video retrieval through example in the paper “MFI-Tree: An effective multi-feature index structure for weighted query application” by Yunfeng He and Junqing Yu. The MFI-Tree employs a tree structure which is beneficial for the browsing application, and retrieves the last level cluster nodes in retrieval application to improve performance. Aggressive decided distance for kNN (ADD-kNN) search algorithm is designed because it can effectively reduce the distance to prune the search space.

According to the different requests of the Web and the heterogeneity of Web servers, Zhang Lin, Li Xiao-Ping and Su Yuan in their paper “A Content-based dynamic load-balancing algorithm for heterogeneous Web server cluster” present a content-based load-balancing algorithm. The mechanism of this algorithm is that a corresponding request is allocated to the server with the lowest load according to the degree of effects on the server and a combination of load state of the server. Furthermore, the algorithm applies a method of random distributing base-probability to assign each request to an appropriate server in terms of their weight.

Increased attention has been given to blind source separation in signal processing, which aims to recover independent sources from their linear instantaneous mixtures without resorting to any prior knowledge. In the paper “Blind separation using second order statistics for non-stationary signals”, Jun Du, Ju Liu, Shengju Sang and Jun Wang propose a new simple BSS technique that exploits second order statistics for non-stationary sources. They use the algebraic structure of the signal model and the subspace structures in order to efficiently recover sources with interference of noise.

Jianfeng Liu, Zhigeng Pan and Xiangcheng Li in their paper “An Accelerometer-Based Gesture Recognition Algorithm and its Application for 3D Interaction” propose an accelerometer-based gesture recognition algorithm. Raw data output by accelerometer should be quantized, and then trained and recognized by discrete Hidden Markov Model. Based on this, they treat gesture as a method of human-computer interaction and use it in 3D interaction subsystem in VR system named VDOM by following steps: establish Gesture-Semantic Map, train standard gestures, finally do recognition.

As a video coding standard, H.264 achieves high compress rate while keeping good fidelity. Haitao Wei, Junqing Yu, and Jiang Li in their paper “The design and evaluation of hierarchical multi-level parallelisms for H.264 encoder on multi-core architecture” propose a hierarchical multi-level parallelisms (HMLP) framework for the H.264 encoder which integrates four level parallelisms – frame-level, slice-level, macroblock-level and data-level – into one implementation. According to the analysis of coding performance on each level parallelism, they propose a method to combine different parallel levels to attain a good compromise between high speedup and low bit-rate.

Ding Ying, Li Wen-Hui, Fan Jing-Tao and Yang Hua-Min present a novel method to robustly and efficiently detect moving objects, even under complex backgrounds, such as those consisting of illumination changes, long shadows, etc. in “Robust moving object detection under complex background”. They integrate the local binary pattern texture measure to extend the moving object detection work for light illumination changing, then introduce HSI color space measure for removing shadows for the background subtraction. Moreover, a novel fuzzy approach using the Choquet integral improves detection accuracy.

Since classification error accumulation in hierarchical methods impacts classification accuracy, in the paper “A novel hierarchical speech emotion recognition method based on improved DDAGSVM”, Qi-Rong Mao and Yong-Zhao Zhan propose a novel hierarchical method based on improved decision directed acyclic graph SVM for speech emotion recognition. The improved DDAGSVM is constructed according to the confusion degrees of emotion pairs. In addition, a geodesic distance-based testing algorithm is proposed for the improved DDAGSVM to give the differently distinguished test samples many decision chances.

Diagnosis strategy is a testing sequence of fault detection and isolation. For the distribution of electronic equipment, Wang Hongxia, Ye Xiaohui and Wang Liang in their paper “Research on optimizing the fault diagnosis strategy of complex electronic equipments” put forward a feasible engineering maintenance method based on the questions of test point selection and diagnosis strategy.

Xiaopeng Wei, Xiaoyong Fang, Qiang Zhang and Dongsheng Zhou propose a new method for matching two 3D point sets of identical cardinality with global similarity but local non-rigid deformations and distribution errors in their paper “3D point pattern matching based on spatial geometric flexibility”. To establish one-to-one identification, they introduce a forward 3D point pattern matching (PPM) method based on spatial geometric flexibility, which considers a non-rigid deformation between the two point-sets.

I would like to give my great thanks to the reviewers for their helpful comments and all of the authors for their contributions, efforts and enthusiasm. Thanks are also due to the ComSIS Consortium, and especially to the Editor-in-Chief of ComSIS, Mirjana Ivanović and other staff in the Editorial Office for their advice and help in making this special issue possible.

Guest Editor
Prof. Qiang Zhang
Key Laboratory of Advanced Design and Intelligent Computing
(Dalian University),
Ministry of Education
Dalian 116622
China
E-mail Address: zhangq@dlu.edu.cn

An Extended Photometric Stereo Algorithm for Recovering Specular Object Shape and Its Reflectance Properties

Zuoyong Zheng¹, Lizhuang Ma¹, Zhong Li², and Zihua Chen^{3,4}

¹Department of Computer Science, Shanghai Jiaotong University, China
zuoyong.zheng@gmail.com, ma-lz@cs.sjtu.edu.cn

²Department of Applied Mathematics, Zhejiang Sci-Tech University, China
lizhongzju@hotmail.com

³Department of Computer Science, East China University of Sci-Tech, China

⁴State Key Laboratory for Novel Software Technology, Nanjing University, China
czh@ecust.edu.cn

Abstract. In Photometric stereo, the existence of specularities hampers to recover the normal map. To deal with this common reflective phenomenon, we introduce a novel representation for specular reflection with a set of specular basis functions with different roughness values. This representation is suitable for any intensively or weakly specular object, and is introduced into the photometric stereo algorithm to recover both the surface shape and its reflectance properties. The reconstructed shapes and re-rendered images validate the proposed algorithm.

Keywords: Photometric Stereo, Specular Reflection, Surface Reconstruction.

1. Introduction

Photometric stereo is an important research field in computer vision, which is used to recover the shape of a static object under varying illuminations. Its first application was proposed by Woodham [1] in 1980's. The underlying principle is very simple: the normal at a surface point can be deduced from its intensity values caused by illuminations from different directions. The surface reflection is assumed to coincide with Lambertian diffuse model and the light source to be parallel. Obviously, pixel intensities are dependent on unknown surface orientations and known lighting directions. The photometric algorithm has three advantages: (1) no surface smoothness is assumed; (2) only multiple light sources are needed when implemented; (3) diffuse parameters can also be obtained. Due to these benefits, photometric stereo is widely used to reconstruct surface shape.

In traditional photometric stereo algorithms, ideal imaging conditions are generally assumed, i.e., Lambertian surface, distant light sources and orthographic camera projection. These constraints, especially the first, limit

the practical use of photometric stereo. The specularities exhibited by most real world objects violate the linear relationship between the pixel intensities and the normals, therefore make the solution deviate from the ground truth. The method coping with specular reflection can be classified into two categories. The first one is to identify and remove specularities using image-base techniques, while in the other group, specular pixels are directly used in analytic or discrete BRDF models. Due to the close coupling of surface shape and its reflectance properties, they are generally worked out simultaneously.

In this paper, an extended photometric stereo algorithm is proposed to recover the shape of a surfaces with specularities, and moreover, to recover its reflectance properties, thus falling into the second category dealing with specular reflection. The distinct feature of this method (also the main contribution of this paper) is to represent any possible specular reflection as a linear combination of a set of specular terms with preset roughness values. The actual specular reflection is projected onto such group of specular basis functions, and the resulting coefficients can be considered to represent the surface reflectance properties. Thus, all diffuse and specular pixels are utilized to recover both the shape and reflectance properties.

2. Previous Works

There are a lot of works to handle specular reflection in photometric stereo. Ikeuchi [2] first applied photometric stereo to specular surfaces. He used three extended light sources and then the reflectance maps for each source in the form of lookup tables. Nayar et al [3] used evenly distributed extended light source to identify accurately the diffuse and specular reflection, and used a so-called "photometric sampling" technique to calculate the surface orientation as well as the component of two kind of reflection. Kay and Caelli [4] used the simplified Torrance-Sparrow model [5] to describe reflectance properties, and introduced the simulated annealing algorithm into photometric stereo to simultaneously recover the normal map and the roughness parameter. Georghiadis [6] used Torrance-Sparrow model to eliminate the ambiguity inherent in uncalibrated photometric stereo for diffuse objects. He established a complicated objective function incorporating all unknowns (reflection coefficients, surface normals, as well as light source intensities and positions). These unknowns are solved by an iterative non-linear optimization process. Shen et al [7] firstly recovered the specular reflectance parameters of the surfaces by a novel optimization procedure, then these parameters are used to estimate the diffuse reflectance and surface normal for each point. Chung and Jia [8] used the Ward model [9] to represent the reflectance properties, and calculated for shadow points their normals and BRDF parameters by using cast shadow information. These parameters are then taken into the Ward model for robustly estimating other points' normals and parameters using an iterative optimization.

Different from the aforementioned methods using analytic BRDF models, other works employed discrete, or data-driven models to replace non-linear BRDFs. Hertzmann and Seitz [10] proposed a smart and practical example-based solution to express reflectance properties, i.e., a purely diffuse ball and a snooker ball to respectively represent the diffuse and specular reflection. In order to calculate the normal at a surface point, all possible points on the two reference balls are searched to approximate its intensity using the linear combination of the two reference points' intensities in least square sense. It is essentially a finite search on reference spheres, and no requirement for geometric and radiometric calibration for the camera and the light sources is its merit.

Alldrin et al [11] firstly analyzed the property of isotropic BRDF models, then express it as a bivariate discrete function in a closed form. After assuming that object materials are made of several "fundamental" materials, they constituted an iterative scheme to simultaneously solve the discrete BRDF, the material weight map and surface normals. Comparing Hertzmann and Alldrin's works, we can find they are similar in spirit to each other, i.e., the idea of "finite fundamental materials" to represent any actual materials.

On the one hand, the existence of specularities makes shape recovery involved in time-consuming non-linear optimization; on the other hand, specularities strongly imply the symmetry between the light direction and view direction with respect to surface normals. Chen et al [12] proposed a photometric algorithm only using specularities. They firstly used the intensity histogram from multiple video frames to judge specular pixels, then computed the normals by using the calculated light position corresponding to that frame resulting in specular pixels. Because specularities are high frequency components in an image, a larger number of images are needed to judge their existence. To exceed this limitation, Francken et al [13] used a LCD monitor displaying binary black-white strip pattern as an array of light sources, instead of single point light source in Chen's work. Usage of such kind of coded light actually increases the light sampling rate, as a result, the required images are cut down.

In the remainder of this paper, we will describe and implement our extended photometric stereo algorithm, which belongs to the discrete-BRDF-model class. Our method is close in principle to Hertzmann's [10] and Alldrin's [11]. In comparison with the former, we used virtual reference objects (instead of actual ones), and express any specular reflection using a few fixed specular basis functions, therefore breaking the constraint that uses reference objects of nearly same specular appearance as the test objects. Compared to the latter, we estimate only surface normals and reflectance coefficients (incorporating their reflectance coefficients and material weight map), while ignoring the roughness estimation because fundamental materials is fixed in advance.

Our algorithm is outlined as follows:

- (1) Render images of virtual reference spheres corresponding to selected specular basis functions;

- (2) For each test object point, carry out a search on the reference spheres to approximate its intensity using reference sphere points' intensities in least square sense. The resulting normal and diffuse parameters are considered as accurate;
- (3) Cluster surface materials in terms of estimated diffuse parameters;
- (4) For each class of material, compute its specular coefficients according to an overdetermined linear equation system with inequation constraints while keeping each point's diffuse parameter fixed.
- (5) Render target objects' images under any illuminations and viewpoints.

The following assumptions are made: (1) isotropic reflection; (2) orthographic camera projection; (3) a single distant light source. The geometrical calibration of the light source (i.e., computing its directions) is needed, but the radiometric calibration (i.e., computing its absolute intensity) is not.

3. Principle

To express complex appearance of real world objects, Hertzmann and Seitz [10] suggest that all materials could be represent as a linear combination of a few fundamental materials. The mathematical expression is as follows:

$$I_p(\mathbf{L}) = \sum_j \lambda_{j,p} f_j(\mathbf{n}_p, \mathbf{v}, \mathbf{L}) \quad (1)$$

Here, I_p represents the point p 's pixel value, while its parameters \mathbf{L} denotes the light directions. $f_j(\mathbf{n}_p, \mathbf{v}, \mathbf{L})$ is the reflectance model for the j th material (whose core is BRDF). Its parameter \mathbf{n}_p and \mathbf{v} are respectively the point p 's normal and the view direction. $\lambda_{j,p}$ represents the weight coefficient of the j th material. In the context of photometric stereo, \mathbf{L} denotes different light directions, therefore I_p and f_j are all vectorial functions. If we see all f_j as independent basis functions, the pixel intensity is accordingly the consequence of their linear combination.

How to choose the basis functions $\{f_j\}$? We start from the Ward BRDF mode [9] which comprises a diffuse and a specular term:

$$BRDF_{Ward}(\mathbf{n}_p, \mathbf{v}, \mathbf{l}) = BRDF_{Ward}(\theta_i, \theta_o, \sigma) = \frac{\rho_d}{\pi} + \frac{\rho_s}{\sqrt{\cos \theta_i \cos \theta_o}} \frac{\exp[-\tan^2 \sigma / \beta^2]}{4\pi\beta^2} \quad (2)$$

The parameters in Ward model are classified into two groups. The first group includes geometrical parameters θ_i , θ_o and σ , denoting respectively the angle between \mathbf{l} (incident direction) and \mathbf{n}_p , \mathbf{v} and \mathbf{n}_p , as well as the angle between the halfway vector \mathbf{h} and \mathbf{n}_p (Fig. 1). The second group contains reflectance parameters ρ_d , ρ_s and β , representing respectively the diffuse and specular coefficient, as well as the roughness.

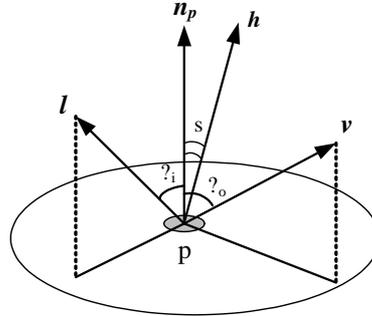


Fig. 1. Geometrical parameters in Ward BRDF

Taking into account the attenuation factor, the reflectance function f can be expressed as follows by means of the Ward model:

$$f(n_p, v, l) = f(\theta_i, \theta_o, \sigma) = \frac{L_0}{r^2} \cos \theta_i \text{BRDF}_{\text{Ward}}(\theta_i, \theta_o, \sigma) \quad (3)$$

Here, L_0 is the absolute light intensity, and r is the distance of target object from the light source. Because the parallel light is assumed, r can be thought of as a constant. Let $\alpha_d = L_0 \rho_d / (r^2 \pi)$, $\alpha_s = L_0 \rho_s / r^2$, f is then expanded as:

$$f(\theta_i, \theta_o, \sigma) = \alpha_d \cos \theta_i + \alpha_s \sqrt{\frac{\cos \theta_i}{\cos \theta_o}} \frac{\exp[-\tan^2 \sigma / \beta^2]}{4\pi\beta^2} \quad (4)$$

where the diffuse coefficient ρ_d and the specular coefficient ρ_s are replaced with α_d and α_s respectively, up to a factor $L_0 / (r^2 \pi)$ and L_0 / r^2 . For the specular term in formula (4), let

$$G(\sigma, \beta) = \frac{\exp[-\tan^2 \sigma / \beta^2]}{4\pi\beta^2} \quad (5)$$

$$e(\theta_i, \theta_o) = \sqrt{\frac{\cos \theta_i}{\cos \theta_o}} \quad (6)$$

and approximate $G(\sigma, \beta)$ with the linear combination of several Gaussians in the same form:

$$G(\sigma, \beta) \approx \sum_{j=1} \omega_j G(\sigma, \beta_j) \quad (7)$$

then substitute (6) and (7) into (4), we get

$$f(\theta_i, \theta_o, \sigma) \approx \alpha_d \cos \theta_i + \sum_{j=1} \alpha_s \omega_j e(\theta_i, \theta_o) G(\sigma, \beta_j) \quad (8)$$

From the above equation, we can see the reflectance function with a Ward model is transformed into a linear combination of a diffuse term and several specular terms with varying roughness. Comparing equation (8) and (1), we

can also find that they are equivalent in essence to each other (let $\lambda_{1,p} = \alpha_d$, $f_1(n_p, v, L) = \cos\theta_i$, at the same time let $\lambda_{j+1,p} = \alpha_s \omega_j$, $f_{j+1}(n_p, v, L) = e(\theta_i, \theta_o)G(\sigma, \beta_j)$, $j=1, 2, \dots$). In other words, all materials on target objects have been represented as a linear combination of a diffuse material and several "pure" specular materials.

Multiple specular materials differ only in their roughness values. In practice, we use $\mathbf{B}=\{\beta_j\}=\{0.800, 0.640, 0.512, 0.410, 0.328, 0.262, 0.210, 0.168, 0.134, 0.107, 0.086, 0.069\}$ to cover any possible roughness. The following figure exhibits the appearances of spheres composed of these specular materials illuminated in a specified light direction (these spheres is called Virtual Reference Spheres, VRS):

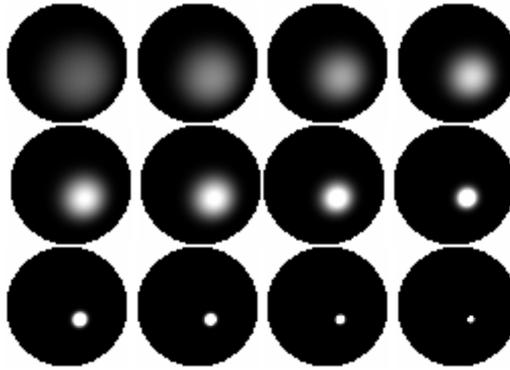


Fig. 2. Twelve rendered specular spheres illuminated in the light direction $[0.4082, 0.4082, -0.8165]^T$. Their roughness correspond the preset roughness collection $\{\beta_j\}$.

The following two figures demonstrate that the Gaussian function $G(\sigma, \beta)$ can be approximated by a collection of other Gaussian functions with various roughness values:

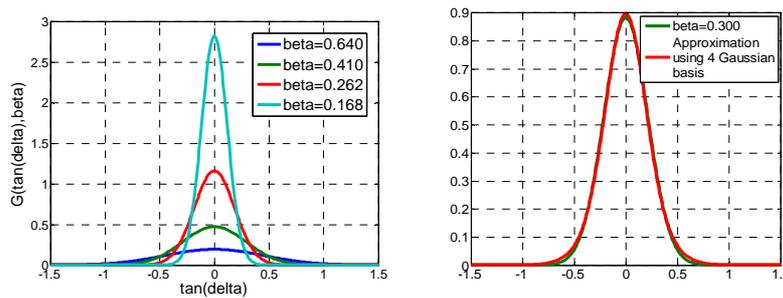


Fig. 3. Left: four Gaussian functions used as basis functions. $\tan(\sigma)$ is used as the independent variable (see formula (5)) Right: the Gaussian function with roughness equal to 0.3000 is approximated by other four Gaussian functions. Their coefficients $\omega_1, \omega_2, \omega_3$ and ω_4 are respectively 0.0003, 0.4843, 0.5768 and 0.0002.

3.1. Recovering the shape and diffuse parameter

After completing the representation for material appearances, estimating the normal per point becomes a search process on all virtual reference spheres. It is detailed as follows:

- (1) Calculate all light directions $\mathbf{L}=\{\mathbf{l}_k\}$ in terms of the specular peak on a shiny steel ball in all test objects images;
- (2) Render the image of the diffuse sphere ($f_1(n,v,l_k)$), and all images of specular spheres ($\{f_j(n,v,l_k), j=2,3,\dots\}$) according to \mathbf{L} . The light intensity L_0 is a preset positive integer (we use 500 in this paper);
- (3) Take a object point p ;
- (4) Search all points on reference spheres for \mathbf{n}_p and $\{\lambda_{j,p}\}$, which minimizes the following objective function:

$$\left\| I_p(L) - \sum_{j=1} \lambda_{j,p} f_j(n_p, v, L) \right\|^2 \quad (9)$$

For color images, equation (9) is actually the quantity accumulated on RGB channel. Also, $\lambda_{j,p}$ is a RGB vector denoting three reflectance coefficients on color channels. The resulting \mathbf{n}_p and $\lambda_{1,p}$ are respectively the accurate normal and the diffuse parameter, while the specular parameters are ambiguous. In implementation, we discarded several smallest pixel values for each point to deal with its potential shading (attached shadow or cast shadow);

- (5) Repeat step (3)-(4), until all object points have been processed;
- (6) Integrate the resulting normal field to construct the triangular mesh.

Executing step (4) is equivalent to solving a overdetermined linear equation system, fulfilled by a matrix pseudo-inverse operation. The resulting \mathbf{n}_p and $\lambda_{1,p}$ are credible, while $\{\lambda_{j,p}, j=2,3,\dots\}$ might be not because for quite a few of points they could not exhibit specular highlights at all under all illumination directions.

3.2. Estimating the specular parameters

In order to estimate specular coefficients reliably, we have made the following assumption (which is not an unreasonable one): points with the same or close diffuse coefficient are thought of as possessing the same material, therefore belonging to a same material category. Thus, estimation of specular parameters for each point evolves to that for each category of material. The diffuse parameter estimated before is naturally chosen as the evidence for clustering materials. In our experiments, the number of categories is pre-defined by observation, then a K-means algorithm is applied to all object points in terms of their diffuse parameter λ_1 .

For each category m , solve its specular coefficients $\{\lambda_j, j=2,3,\dots\}_m$ by minimizing the following objective function while holding the diffuse parameter λ_1 for each point:

$$\sum_{\forall \text{ pixel } p \in \text{class } m} \left\| I_p(L) - \sum_{j=1} \lambda_j f_j(n_p, v, L) \right\|^2 \quad \text{s.t.} \quad \sum_{j=1} \lambda_j f_j(n_p, v, l_k) \geq 255, \forall I_p(l_k) = 255 \quad (10)$$

The additional constraint in function (10) is used to simulate the saturation effect of a digital camera, i.e. the pixel value will be truncated to a maximum (usually 255) when exposure on that CCD cell location is beyond a physical threshold. The above optimization is actually a linear least square problem with inequation constraints. When implemented, the *lsqlin* function of Matlab is used.

4. Experimental Results

We have tested our algorithm on five real objects, named respectively as Bear, Horse, Rabbit, Heart and Peanut. The light source is a hand-held torch, illuminating these objects from 15-19 directions, and the same number of images was captured by a Canon EOS 5D digital camera. This camera was fixed on a tripod with its optic axis being roughly perpendicular to the base plane of the test objects. The capture was accomplished by an accessory remote-control switch attached to the camera, and the images were uploaded onto a computer connected to the camera via a USB line. The test objects were separated from the background by using edge detection, then were applied the algorithm depicted in section 2. Finally we obtained their surface shapes and reflectance parameters.

The recovering algorithm was written in Matlab code, exporting resultant triangular meshes with reflection parameters for each vertex, which were then imported into a rendering program written in Visual C++ 6.0 and OpenGL 1.2 for display.

Among the five test objects, all but Peanut have specular reflection. From these experimental results, we can see objects shapes are well recovered, and the rendered images are perceptually similar to those original ones, which validate the recovered reflectance parameters (see Fig. 4 and Table 1).

Table 1. The recovering time and relative mean error for five test objects. The relative mean error refers to intensity differences between original and rendered pixels

Object name	Bear	Horse	Rabbit	Heart	Peanut
Recovering time (min.)	45	29	24	17	23
Relative mean error	8.84%	8.02%	4.21%	7.53%	5.29%

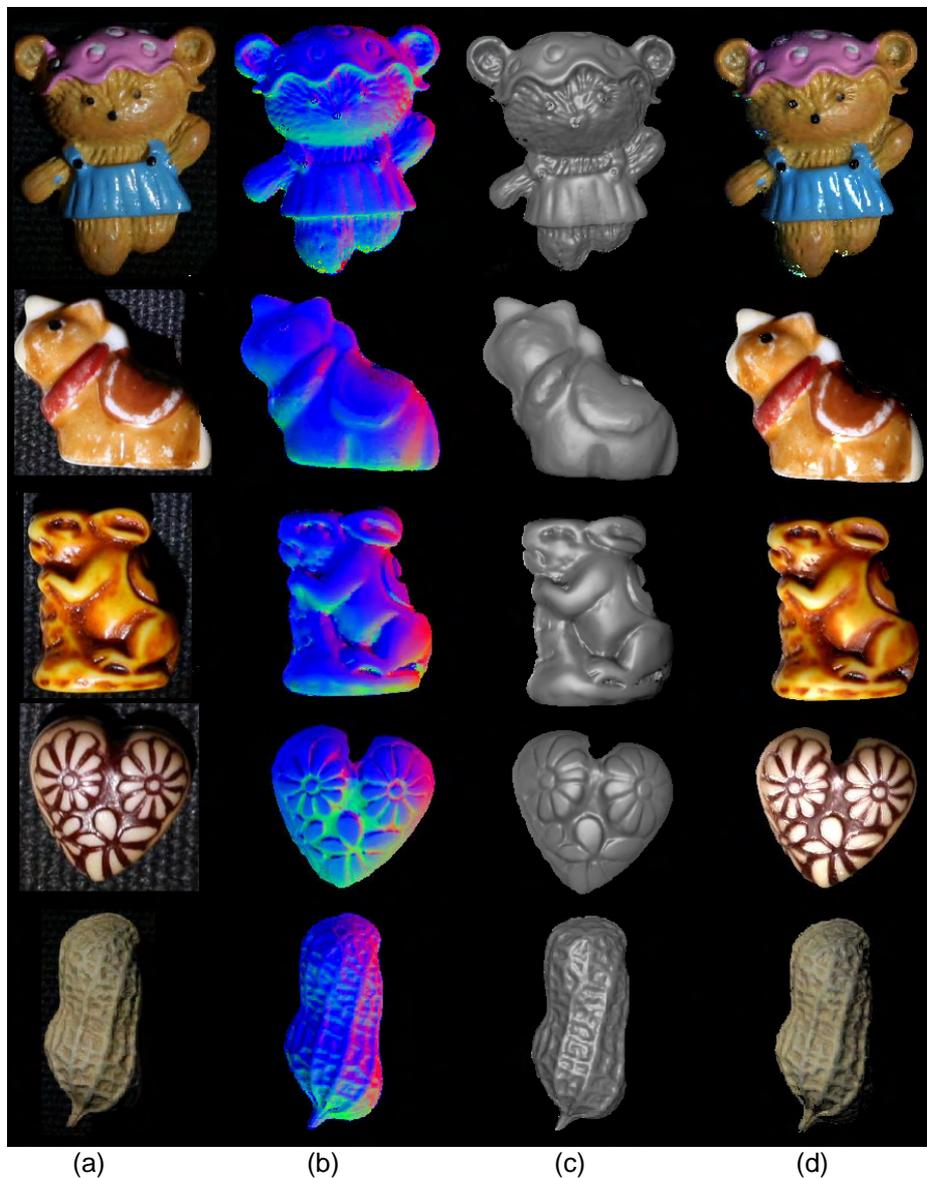


Fig. 4. (a) original test objects images (b) recovered normal maps, the x , y and z component of a normal are respectively encoded in RGB channels (c) reconstructed triangular meshes (d) images rendered according to the recovered shapes and reflection parameters, visually resembling those original ones

We also give rendered images of some objects under novel viewpoints. Because both their shapes and reflectance properties are in hand, the visual appearances under novel viewpoints and light directions can be predicted.



Fig. 5. The left Bear's shape was obtained using the traditional photometric algorithm, while the right shape was obtained using our extended algorithm



Fig. 6. Rendered images for some objects under a novel viewpoint

5. Conclusion

In this paper we proposed an extended photometric stereo algorithm for recover both shape and reflectance properties of a specular object without any removal of specularities. In terms of the recovered shape and reflection parameters, the realistic rendering can be achieved. This algorithm can handle any isotropic reflection, no matter it contains intensively or weakly specular highlight. Slow computation is its main drawback, therefore our future work is to determine the roughness range using a Laplacian-based technique, and further, to reduce the number of basis functions.

6. Acknowledgement

This work was supported by National Basic Research Program of China (973 Program) under Grant No. 2006CB303105, National Natural Science Foundation of China under Grant No. 60873136, Fund for PhD Training under

Grant No. 20050248046, and the Open Project Program of the State Key Lab for Novel Software Technology in Nanjing University. We especially thank Xuezhong Xiao, Na Chu and Yan Wang for their help during the preparation of this paper.

7. References

1. Woodham, R.: Photometric Method for Determining Surface Orientation from Multiple Images. *Optical Engineering*, Vol. 19, No. 1, 139-144. (1980)
2. Keuchi, K.: Determining Surface Orientations of Specular Surfaces by Using the Photometric Stereo Method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 3, No. 6, 661-669. (1981)
3. Nayar, S.K., Ikeuchi, K., Kanade, T.: Determining Shape and Reflectance of Hybrid Surfaces by Photometric Sampling. *IEEE Transactions on Robotics and Automation*, Vol. 6, No. 4, 418-431. (1990)
4. Kay, G., Caelli, T.: Estimating the Parameters of An Illumination Model Using Photometric Stereo. *Graphical Models and Image Processing*, Vol. 57, No. 5, 365-388. (1995)
5. Torrance, K.E., Sparrow, E.M.: Theory for Off-specular Reflection from Roughened Surfaces. *Journal of the Optical Society of America*, Vol. 57, No. 9, 1105-1114. (1967)
6. Georgiades, A.S.: Incorporating the Torrance and Sparrow Model of Reflectance in Uncalibrated Photometric Stereo. In *Proceedings of the 9th International Conference on Computer Vision*. IEEE Computer Society Press, Nice, France, 816-823. (2003)
7. Shen, L., Machida, T., Takemura, H.: Efficient Photometric Stereo Technique for Three-dimensional Surfaces with Unknown BRDF. In *Proceedings of the 5th International Conference on Digital Imaging and Modeling*. IEEE Computer Society Press, Ottawa, Canada, 326-333. (2005)
8. Chung, H.S., Jia, J.: Efficient Photometric Stereo on Glossy Surfaces with Wide Specular Lobes. In *Proceedings of 18th International Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society Press, Santa Barbara, US, 1-8. (1998)
9. Ward, G.: Measuring and Modeling Anisotropic Reflection. In *Proceedings of 21st Annual Conference on Computer Graphics and Interactive Techniques*. Addison-Wesley, Orlando, US, 239-246. (1994)
10. Hertzmann, A., Seitz, S.M.: Example-based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 8, 1254-1264. (2005)
11. Alldrin, N., Zickler, T., Kriegman, D.: Photometric Stereo with Non-parametric and Spatially-varying Reflectance. In *Proceedings of 28th International Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society Press, Anchorage, US, 1-8. (2008)
12. Chen, T.B., Goesele, M., Seidel, H.P.: Mesostructure from Specularity. In *Proceedings of 26th International Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society Press, New York, US, 825-832. (2006)
13. Francken, Y., Cuypers, T., Mertens, T., Gielis, J., Bekaert, P.: High Quality Mesostructure Acquisition Using Specularities. In *Proceedings of 28th International Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society Press, Anchorage, US, 1-7. (2008).

Zuoyong Zheng, Lizhuang Ma, Zhong Li, and Zhihua Chen

Zuoyong Zheng is currently a PhD student at the Department of Computer Science, Shanghai Jiaotong University, China. He has participated in several open projects supported by national level scientific research funds. His research interests include shape modeling and realistic rendering

Lizhuang Ma is a professor of the Department of Computer Science, Shanghai Jiaotong University, China. His research areas are computer graphics and computer aided geometry design. He has (co-)authored around 160 research papers, and has been the chair and/or member of program committees of many international conferences.

Zhong Li is an associate professor of mathematics and science at Zhejiang Sci-Tech University. He received the Ph.D. degree in mathematics from Zhejiang University in 2003. He was a post-doctoral fellow of computer science at Shanghai Jiao Tong University from 2004 to 2006 and was a visiting scholar of computer science division at University of California at Berkeley, U.S.A. from 2007 to 2008. His current research interests are mesh processing and garment CAD.

Zhihua Chen is a lecturer in the School of Information and Engineering at East China University of Science and Technology. So far, he has authored over 20 research papers. He is a member of CCF(China Computer Federation). His research interests include CG, CAD and Compute Animation.

Received: May 02, 2009; Accepted: July 12, 2009.

Extracting PCB components based on color distribution of highlight areas

Zhou Zeng¹, LiZhuang Ma¹, and ZuoYong Zheng¹

¹ The School of Electronic, Information and Electrical Engineering,
the Shanghai Jiao Tong University, MinHang Campus,
NO.800 Dong Chuan Road, Shanghai, 200240, P.R.China
bluecourse@china.com, ma-lz@cs.sjtu.edu.cn, Zuoyong.Zheng@gmail.com

Abstract. This paper investigates methodologies for locating and identifying the components on a printed circuit board (PCB) used for surface mount device inspection. The proposed scheme consists of two stages: solder joint extraction and protective coating extraction. Solder joints are extracted by first detecting all the highlight areas, and then recognizing and removing the invalid highlight areas which are mainly markings and via-holes. We sum up three color distribution features. And the invalid highlight areas are recognized and removed by comparing the features of the target objects and the reference objects. The sequence of color distribution as a new clue has been applied to clustering solder joints. Each protective coating is extracted by the positions of the clustered solder joints. Experimental results show that the proposed method can extract most of components effectively.

Keywords: machine vision, component extraction, solder joints, protective coatings.

1. Introduction

With the fast development of surface mounting technology, the need for automatic inspection has been ever increasing. A portion of PCB surface defect inspection involves inspecting the solder joints on the PCB and many different inspection approaches have been developed (Teoh et al. [13], Bartlet et al. [1], Kim et al. [6] [7]). Because of the complexity of the board surface, the identification of solder joint locations is relatively difficult. Loh et al [3] use a slant map to extract the shape information of a solder joint, which is based on the slant angle of the solder joint surface. But to their approach, there exist a number of factors that can affect the solder joint shape, including the gravity, the solder ability of the solder pad and so on. For these reasons, their approach is limited. The histogram based methods are used extensively to extracting solder joints. Kim et al. [7] take three frames of images which are sequentially captured as three layers of LEDs and turned on, one after another. From these images, soldered regions are segmented by using x/y

projection and thresholding. This approach just works for the local areas where solder joints are concentrated. Wu and Wang [4] present an efficient approach for PCB components detection. The suggested method is based on multiple template matching and a modified species based particle swarm optimization (SPSO) with three specific acceleration strategies. This kind of approach has a high degree of accuracy for a specified object. But it needs to sample lots of templates for a complex PCB, which has many different components.

This paper explores automated object-recognition techniques for extracting multiple objects (e.g. IC, chip resistor, transistor, diode and etc.) in a PCB image. From the perspective of PCB image, a typical chip component consists of protective coating, markings and solder joints. The proposed approach detects solder joints based on their specular attribute and determines specular interval of gray levels by multilevel thresholding algorithm. The markings, via-holes and other invalid specular areas are recognized and removed by comparing color distribution features of them with those of reference object in (u', v') chromaticity coordinates. The sequence of color distribution as a new clue has been applied to clustering solder joints. Each protective coating is extracted by the positions of the clustered solder joints. According to the experimental results, the developed algorithm can effectively recognize most of components without any restrictions.

2. Illuminant System

We capture images by PCB inspection device VT-RNS (Omron Corp.). Three layers of ring-shaped LEDs with the different illumination angles and input camera are controlled by the host computer. The illumination design is similar to the equipment that is used for Kim and Cho [6]. We acquire an image when all LEDs are turned on in order to get specular areas of solder joints as many

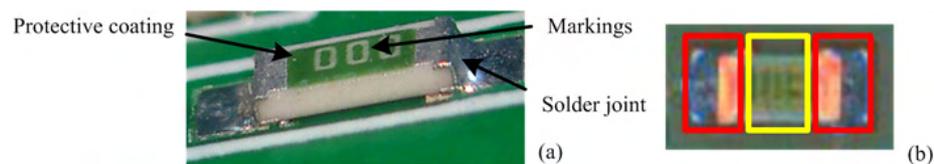


Fig. 1. A typical component. (a) component structure; (b) image of the component under three layers of ring shaped LEDs illuminant system.

as possible. Fig.1 (a) shows a typical chip component. From the perspective of PCB image, it consists of a protective coating, multiple markings and more than two solder joints. Fig.1 (b) shows the projected image under illuminant system. The highlight area of solder joints shows up as the 2D projection of R, B and G three LEDs, and the color distribution of highlight area shows regularity. The proposed algorithm consists of two parts: One is to extract

solder joints. That is, to locate solder joints and to evaluate their ranges (see red rectangles in Fig.1 (b)). This is discussed in section 3. Another is to extract protective coatings. That is, to locate protective coatings and to evaluate their ranges (see a yellow rectangle in Fig.1 (b)). This is discussed in section 4.

3. Extracting solder joints

In general, the algorithm of extracting solder joints is divided into two stages: one is to detect all highlight areas; another is to recognize and remove invalid highlight parts based on the color distribution features. Then the remained highlight areas only contain solder joints. In Sect.3.1, the specular detection algorithm is presented. In Sect.3.2, GMM based color distribution feature extraction and comparison approach is introduced. The implement details of extracting solder joints are described in Sect.3.3.

3.1. Specular detection

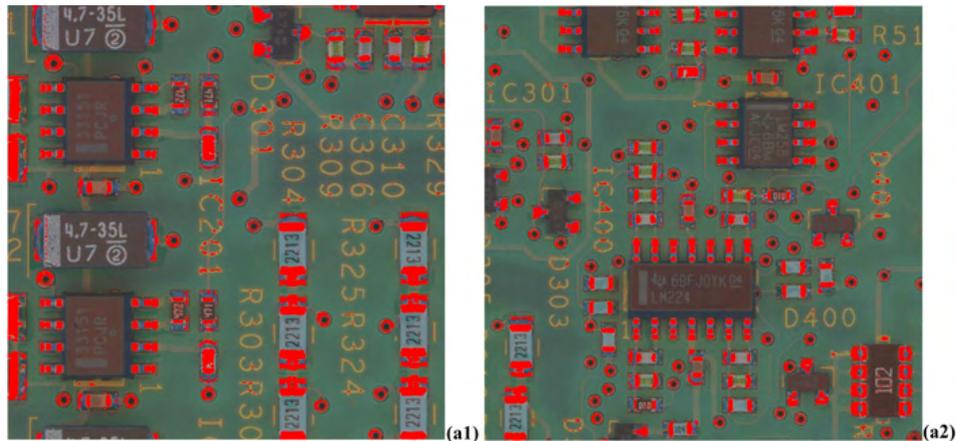


Fig. 2. PCB images and specular detection results. (a1) and (a2) are two PCB images I_a and I_b . Red points are detected highlight pixels.

Many specular detection methods have been introduced in the past decade, such as [7][8][11][12]. Yen et al. [14] have proposed a new criterion for multilevel thresholding. We present a novel approach based on automatic multilevel thresholding for detecting specularity. Implementation details are described as follows: (1) Image preprocess. Decrease brightness and sharpen the contrast of input image by

$$p' = [(p + (b + 100) \cdot 1.27) - 127] \cdot [(c + 100) \cdot 0.01] + 127 \quad (1)$$

where p is a original pixel; brightness coefficient $b \in (-100,100)$ and contrast coefficient $c \in [0,100)$. This operation is favorable for minimizing disturbance from background and makes the specular areas more obvious. (2) Convert color space from RGB to CMYK. CMYK produces a lower level of color details than RGB does. This helps to decrease the discontinuity of specular areas. (3) Construct the histogram of K channel and evaluate the specularity interval by automatic multilevel thresholding approach[14]. By minimizing the cost function, the classification number that the gray-levels of K channel should be classified and the threshold values can be determined automatically. The last classification belongs to specularity. We select two typical PCB images I_a and I_b to illustrate the experiment results. Fig.2 (a1) and (a2) show specular detection results of I_a and I_b respectively.

3.2. GMM based color distribution features extraction and comparison

Color cues have been shown to offer several significant advantages for certain tasks in visual perception, such as [5][9]. The proposed approach also takes color as a cue for recognizing the specified components based on Gaussian Mixture Model (GMM). A framework is developed for recognizing the target objects according to the following procedures: First, project all highlight pixels into $U'V'$ space and model the color distribution of them by GMM. And then, sample several reference objects and all target objects, and extract their color distribution features. Finally, recognize the target object with a comparison between features of the target object and features of some reference objects.

3.2.1. Modeling color distribution of highlight pixels in $U'V'$ space

In order to keep color constant, first, we convert all recognized highlight pixels into LUV space, and then project them into (u',v') chromaticity coordinates with the following transformation

$$u' = u / (13l) + u'_n, v' = v / (13l) + v'_n \quad (2)$$

where l, u, v are LUV channels; the quantities u'_n and v'_n are the (u',v') chromaticity coordinates of a specified white point. In our case, we set $u'_n = 0.200331$ and $v'_n = 0.474959$ with standard illuminant C.

Fig.3 (a) shows the visible gamut in (u',v') perceptually uniform coordinates. The color distribution is modeled based on GMM. First, all highlight pixels are classified into three areas A_b , A_g and A_r . In Fig.3 (a), three black lines segment (u',v') space into above three areas. Each highlight pixel is assigned to the specified area according to the following rule

$$\alpha = \arctan(|v_{\max} - v'_n - v'_i| / |u'_i - u'_n|) \cdot 180 / \pi, P_i \in \begin{cases} A_r, & \alpha \leq 89^\circ \cup \alpha \geq 304^\circ \\ A_g, & \alpha \in (89^\circ, 192^\circ) \\ A_b, & \alpha \in [192^\circ, 304^\circ) \end{cases} \quad (3)$$

where v_{\max} is the maximum value of coordinate axis v' ; chromaticity coordinates (u'_i, v'_i) and (u'_n, v'_n) locate highlight pixel P_i and white point P_w respectively; Angle α between vector $\overline{P_w P_i}$ and horizontal line (see Fig.3 (a)). Fig.3 (b1) and (b2) show color distributions of I_a and I_b .

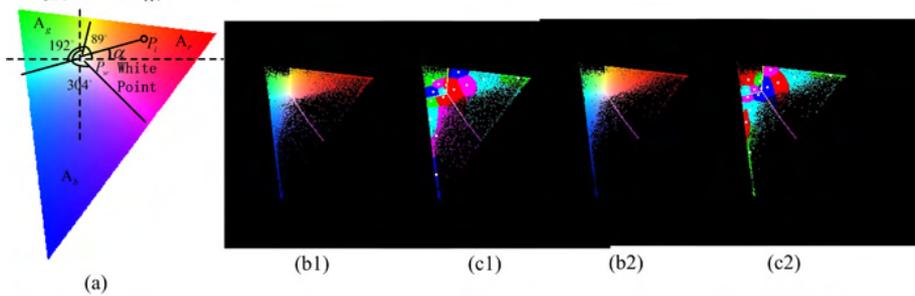


Fig. 3. Color distributions. (a) CIE 1976 uniform chromaticity scale diagram. (b1) and (b2) are color distributions in chromaticity coordinates of I_a and I_b respectively; (c1) and (c2) are evaluated Gaussian mixture distributions with five mixtures of I_a and I_b respectively. Each kind of color represents a mixture.

After that, color distribution of each area is modeled by Gaussian mixture distribution. The EM (Expectation-Maximization) algorithm estimates the parameters of the multivariate probability density function in a form of the Gaussian mixture distribution with a specified number of mixtures (Bishop [2] and Render [10]). Consider the set of the feature vector $\{x_1, x_2, \dots, x_n\}$: n vectors from d -dimensional Euclidean space drawn from a Gaussian mixture

$$\rho(x|a_k, S_k, \pi_k) = \sum_{k=1}^m \pi_k P_k(x), \pi_k \geq 0, \sum_{k=1}^m \pi_k = 1 \quad (4)$$

$$\rho_k(x) = \varphi(x|a_k, S_k) = \frac{1}{(2\pi)^{d/2} |S_k|^{1/2}} \cdot e^{\left\{-\frac{1}{2}(x-a_k)^T S_k^{-1} (x-a_k)\right\}} \quad (5)$$

where m is the number of mixtures; ρ_k is the normal distribution density with the mean a_k and covariance matrix S_k ; and π_k is the weight of the k th mixture. Given the number of mixtures m and the samples $\{x_1, x_2, \dots, x_n\}$ the algorithm finds the maximum-likelihood estimates (MLE) of the all the mixture parameters, i.e., a_k , S_k and π_k . Corresponding to pixels within A_r , A_b and A_g , three groups of GMM parameters, $\{a_{k,r}, S_{k,r}, \pi_{k,r}\}$, $\{a_{k,b}, S_{k,b}, \pi_{k,b}\}$ and $\{a_{k,g}, S_{k,g}, \pi_{k,g}\}$ are evaluated for modeling color distribution of all highlight

pixels. For simplicity, the GMM parameters are denoted as $\{a_{k,c}, S_{k,c}, \pi_{k,c}\}, c \in \{r, g, b\}$. Fig.3 (c1) and (c2) show the modeling results of I_a and I_b respectively with $m = 5$.

3.2.2. Extracting distribution features of the specified object

Three color distribution features of each specified object are evaluated:

(1) Mean distribution probability of each mixture. A $3 \times m$ matrix M_{prob} is defined to store $3m$ mean probabilities, and $M_{prob}(i, j)$ denotes the mean probability of j th mixture in i th area.

(2) Mean distance between pixels in each mixture and white point. A $3 \times m$ matrix M_{dist} is defined to store $3m$ mean distances, and $M_{dist}(i, j)$ denotes the mean distance of j th mixture and in i th area.

(3) The ratios of pixel number of three areas. A vector V_{ratio} is defined to store three ratios, and $V_{ratio}(i)$ denotes the ratio of pixel number of i th area.

Some terms are used to describe the steps of evaluating color distribution features, they are defined as follows: ψ_{tar} and ψ_{ref} denote the color distributions of the target object and the reference object respectively; A_c denotes the area and $c \in \{r, g, b\}$; $\tau_{c,i}$ denotes the i th mixture in A_c ; $N(c, i)$ records pixel number of the i th mixture and in A_c ; $\rho_k(c, i)$ denotes the normal distribution density of pixel P_k in the i th mixture and in A_c ; d_k denotes the distance between P_k and white point.

Suppose that there are n highlight pixels $P = \{P_k | P_1, P_2, \dots, P_n\}$ in an object. In order to evaluate three features, first, the $3 \times m$ matrix M_{dist} , M_{prob} and a vector V_{ratio} are initialized as zero. And then, P_k is assigned to the specified area A_c by (3). For each pixel P_k , computes its m probabilities $\rho_k(c, i)$ by (5) with evaluated parameters $\{a_{i,c}, S_{i,c}, \pi_{i,c}\}, i \in [1, m]$ and computes its distance d_k . Finally, mean probabilities, mean distances and pixel number ratio are evaluated by:

$$N_c = \sum_{i=1}^m N(c, i), c \in \{r, g, b\}, i \in [1, m]$$

$$\text{For all } P_k, \text{ and } P_k \text{ within } A_c, M_{prob}(c, i) = \sum_{k=1}^{N_c} \rho_k(c, i) / N_c \quad (6)$$

$$\text{For all } P_k, \text{ and } P_k \text{ within } \tau_{c,i}, M_{dist}(c, i) = \sum_{k=1}^{N(c,i)} d_k / N(c, i)$$

$$V_{ratio}(c) = N_c / n$$

To summarize, we express differences between the color distributions of two objects in three levels. V_{ratio} describes the difference between areas, M_{dist} describes the difference between mixtures and M_{prob} describes the difference within single mixture. Only by working together, three features can effectively describe the color distribution. It's likely to causing ambiguity just relying on one or two of features.

3.2.3. Recognizing the target object by feature comparison

In order to determine whether ψ_{ref} and ψ_{tar} are of the same type, there are three cases are considered: (1) ψ_{tar} and ψ_{ref} share the same mixtures; (2) All pixels of ψ_{tar} within the subset of the mixtures of ψ_{ref} ; (3) Some pixels of ψ_{tar} are outside the mixtures of ψ_{ref} . The relative difference on mean distribution probability is evaluated according to formula

In case1 and 2: if $M'_{dist}(c, i) > 0$,

$$f_{prob}(c, i) = \frac{|M_{prob}(c, i) - M'_{prob}(c, i)|}{f_{non_zero}(M_{prob}(c, i), M'_{prob}(c, i))}$$

In case3: if $M'_{dist}(c, i) = 0$ and $M_{dist}(c, i) > 0$,

(7)

$$f_{prob}(c, i) = \left(\frac{|M_{prob}(c, i) - M'_{prob}(c, i)|}{f_{non_zero}(M_{prob}(c, i), M'_{prob}(c, i))} \right) \cdot \mu, \mu > 1.0,$$

$$\lambda_{prob}(c) = \left(\sum_{i=1}^m f_{prob}(c, i) \right) / m$$

where M_{prob} , M_{dist} and V_{ratio} are features of ψ_{tar} , M'_{prob} , M'_{dist} and V'_{ratio} are features of ψ_{ref} ; and $\lambda_{prob}(c)$ denotes the relative difference on mean distribution probability in A_c ; function f_{non_zero} is used to select a nonzero variable; function $f_{prob}(c, i)$ denotes a relative difference on mean distribution probability of the i th mixture and in A_c . In general, we sample multiple reference objects simultaneously in order to maintain stability of reference features. It might lead to the distribution scope of ψ_{tar} that is less than or equal to the scope of ψ_{ref} . So case 1 and case 2 are considered normal, but there is a huge variation between ψ_{tar} and ψ_{ref} in case 3. In order to show this variation, a penalty μ is attached to enlarge the relative differences. After that, the relative differences on mean distances is evaluated in three areas according to formula

$$\lambda_{dist}(c) = \left(\sum_{i=1}^m \frac{|M_{dist}(c, i) - M'_{dist}(c, i)|}{f_{non_zero}(M_{dist}(c, i), M'_{dist}(c, i))} \right) / m$$
(8)

where $\lambda_{dist}(c)$ denotes the relative difference on mean distance in A_c . Finally, evaluate the total difference of three features between ψ_{ref} and ψ_{tar} by

$$f_{ratio}(c) = \frac{|V_{ratio}(c) - V'_{ratio}(c)|}{\sum_{c=1}^3 |V_{ratio}(c) - V'_{ratio}(c)|}, \lambda = \sum_{c=1}^3 (\alpha \cdot \lambda_{prob}(c) + (1-\alpha) \cdot \lambda_{dist}(c)) \cdot f_{ratio}(c) \quad (9)$$

where $f_{ratio}(c)$ denotes a relative difference on ratio of pixel number in channel c , λ denotes total difference, coefficient α is used to adjust the ratio between difference of mean probability distribution and difference of mean distance, and $\alpha \in [0,1]$. Usually, we set $\alpha = 0.6$. If $\lambda < 0.5$, then ψ_{ref} and ψ_{tar} are of the same type.

3.3. Implement details of extracting solder joints

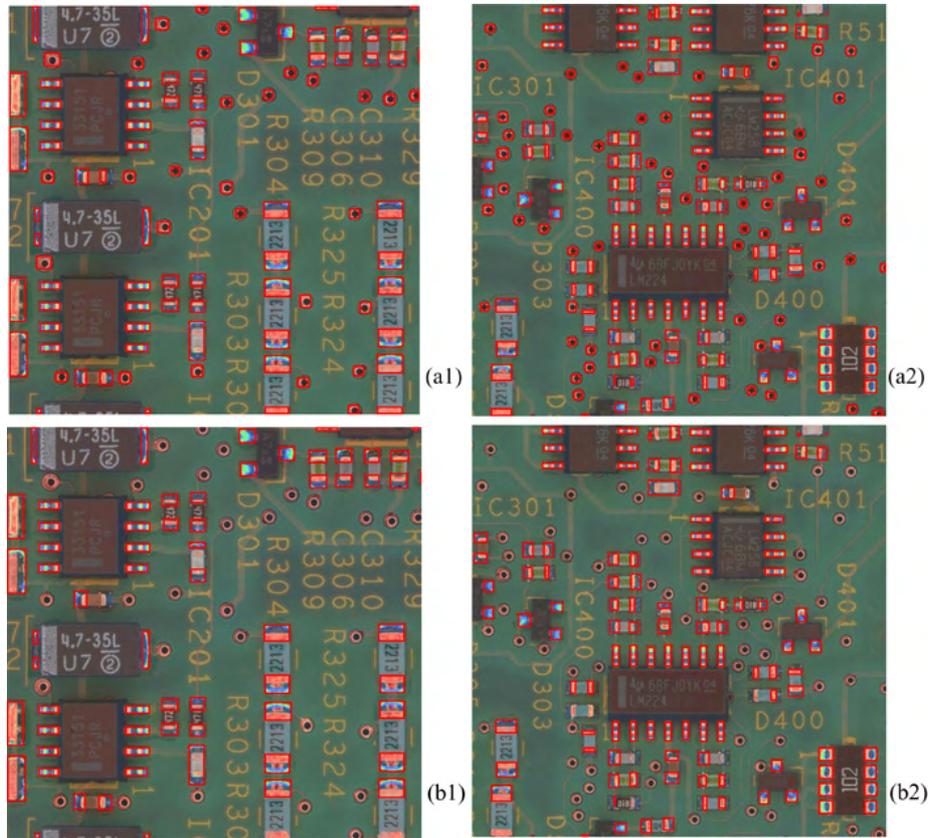


Fig. 4. Recognize and remove results of PCB images I_a and I_b . (a1) and (a2) show the results after removing the markings and merging the remaining clusters; (b1) and (b2) show the results after removing via-holes.

The five steps of extracting solder joints are described as follows:

First, project all highlight pixels in $u'v'$ chromaticity coordinate and model their color distribution based on GMM (refer to Sect. 3.2.1).

Second, cluster all highlight pixels. We traverse all highlight pixels and classify them into many clusters according to their differences in K channel of CMYK color space by eight neighborhood search. We define c is a single cluster of cluster set $\{C\}$.

Third, evaluate features M_{prob} , M_{dist} and V_{ratio} of each cluster c by (6).

Forth, remove markings. Sample multiple typical markings as a reference object ψ_{ref} , and evaluate distribution features M'_{prob} , M'_{dist} and V'_{ratio} of ψ_{ref} by (6). Make each cluster c as the target object ψ_{tar} and evaluate total difference λ between ψ_{ref} and ψ_{tar} by (7), (8) and (9). If $\lambda < 0.5$, then c is a marking and remove it from set $\{C\}$.

Fifth, merge the remaining clusters of $\{C\}$ according to their distances. The specular areas of a solder joint may be not continuous because of the changes in surface normal and light source position. It makes the highlight pixels of a solder joint are usually classified into multiple clusters, thus leads to recognition errors. We alleviate this problem by merging adjacent clusters before removing via-holes.

Finally, remove via-holes. Sample some typical via-holes as the reference object, and then remove via-holes from $\{C\}$ in the same way as step forth. Now the remaining clusters of $\{C\}$ only contain solder joints.

Fig.4 shows solder joints extraction results of PCB images I_a and I_b . In Fig.4 (a1) and (a2), all markings of I_a and I_b have been removed and the merging clusters have been merged. Fig.4 (b1) and (b2) show the results that via-holes have been removed and the remaining clusters of $\{C\}$ only contain solder joints.

4. Clustering solder joints and extracting their protective coatings

We have found that the color distribution of solder joints shows regularity. The regularity can be summed up as follows: for each solder joint, the color distribution of a part or a whole highlight region of a solder joint is changed in a sequence of {R, B, G} in the direction from protective coating to solder joint. According to this principle, we can determine the direction from solder joint to the protective coating that it is connected to. The determined directions offer important clues for extracting protective coatings. The main procedure is as follows: first, determine the direction of each solder joint based on above regularity (see section 4.1). Second, cluster all solder joints of each component by directions of each component and the gray level interval of PCB's background. Finally, the protective coating of each component is

extracted according to the positions of all connected solder joints (see section 4.2).

4.1. Determining the direction of each solder joint

The components mount to the PCB along a horizontal (U axis) or vertical (V axis) direction. So, first of all, we determine the axis of color distribution. And then, we determine the direction of axis.

4.1.1. Determining the axis of color distribution

The color distribution is changing with its axis. Therefore, the dispersion of color distribution along with axis is greater than the dispersion along a line perpendicular to the axis. We define the dispersion by the coefficient of variance. Suppose that the size of a solder joint is $w \times h$, and γ_u and γ_v are defined as the mean coefficients of variance of data that sample along with the U and V axes respectively. Symbol ϕ_{axis} denotes the direction of axis. In order to determine ϕ_{axis} , first, we sample h rows of data separately from U and V channels along with the U axis. The coefficient of variance of each row of data is evaluated by (10) and stored into arrays ρ_{uL} and ρ_{uU} respectively, and ρ_{uL} and ρ_{uU} with the length h .

$$\mu = \left(\sqrt{\sum_{i=1}^n (D(i) - \bar{D})^2 / n} \right) / \bar{D} \quad (10)$$

where μ denotes coefficient of variance; D and \bar{D} denote the sample data and its mean value respectively; n expresses the number of sample data. After that, we sample w columns of data separately from U and V channels along with the V axis. The coefficient of variance of each column of data is evaluated by (10) and stored into arrays ρ_{vL} and ρ_{vU} respectively with the length w . γ_u and γ_v are evaluated by

$$\gamma_u = \sum_{i=1}^h \sqrt{\rho_{uL}(i)^2 + \rho_{uU}(i)^2} / h, \quad \gamma_v = \sum_{i=1}^w \sqrt{\rho_{vL}(i)^2 + \rho_{vU}(i)^2} / w \quad (11)$$

If $\gamma_u \geq \gamma_v$, $\phi_{axis} = 0$, else $\phi_{axis} = 1$. Where 0 and 1 denote the color distribution along with the U and V axes respectively.

4.1.2. Determining the direction of axis

According to the color distribution regularity mentioned above, we define the direction of axis by four steps: First of all, sample color distribution data along with the determined axis; Then, filter the useless data; After that, select some

valid data from the sample data along with the forward and reverse directions of axis respectively, and fit them by Dose-response curve; Finally, determine the direction of axis by analyzing the fitting parameters. Implementation details are described as follows:

Step 1: sample color distribution data. We decrease brightness and sharpen the contrast of original PCB image by (1) for enhancing the features of color distribution. At the same time, channels R, B and G are encoded in integer 3, 2, and 1 respectively. For each pixel included in $w \times h$ region, we compare its three channels and store its channel code with maximal value into a $w \times h$ matrix M_p . Then, the color distribution data is evaluated and stored in a vector V_{mean} . If $\phi_{axis} = 0$, then the length of V_{mean} is w ; otherwise, the length of V_{mean} is h . Color distribution data is evaluated by the following rules:

$$\begin{aligned} \text{If } \phi_{axis} = 0, \text{ for } i = 1 \text{ to } w, V_{mean}(i) &= \sum_{j=1}^h M_p(i, j) / h, \\ \text{If } \phi_{axis} = 1, \text{ for } i = 1 \text{ to } h, V_{mean}(i) &= \sum_{j=1}^w M_p^T(i, j) / w \end{aligned} \quad (12)$$

Fig.5 (a1)-(a3) shows three typical solder joints. In Fig.5 (b1)-(b3), the black curve shows the color distribution data.

Step 2: filter the useless data. In some cases, the color distributions of some solder joints become more complex and make their color distribution curves not smooth enough, because of the variances in surface normal and angle of illumination incidence. It may lead fitting to a failure in the next step. Therefore, we remove some useless data in order to smooth the curve effectively. First, we map each data $V_{mean}(i)$ to three intervals based on the following rules and save it into $V'_{mean}(i)$.

$$V'_{mean}(i) = \begin{cases} 1, & \text{if } V_{mean}(i) \in [1, 1.5) \\ 2, & \text{if } V_{mean}(i) \in [1.5, 2.3) \\ 3, & \text{if } V_{mean}(i) \in [2.3, 3.0] \end{cases} \quad (13)$$

In Fig.5 (b1) to (b3), the red dotted line shows the mapped data. Furthermore, we check the contiguous data with the same value in V'_{mean} . If the scale of contiguous data is less than 2, this kind of data is useless and ought to be removed. For example, in Fig.5, the 11th, 12th and 48th data in (b3) are useless. They are enclosed with blue ellipses.

Step 3: fit color distribution data. The color distribution curve can be fitted by some growth models, such as dose-response function. We define S_f and S_r as two data segments. Check data along with the forward direction of axis, if $V'_{mean}(i) \geq V'_{mean}(i+1)$, add $V'_{mean}(i)$ into S_f . Check data along with the reverse direction of axis, if $V'_{mean}(i) \geq V'_{mean}(i-1)$, add $V'_{mean}(i)$ into S_r . Data segment with length is less than four is invalid because dose-response function has four parameters. In Fig.5 (b1)-(b3), the picked data segments are marked by the

green lines. Where S_f of (b2) is invalid. After that, we fit S_f and S_r by dose-response function. A standard dose-response function is defined by four parameters

$$y = A_1 + (A_2 - A_1) / (1 + 10^{(LOGx0-x)p}) \quad (14)$$

where A_1 is the baseline response (Bottom); A_2 is the maximum response (Top); p is the slope and $LOGx0$ is the drug concentration that provokes a response halfway between baseline and maximum. In Fig.5 (c1) to (c5), the red curve shows fitting result.

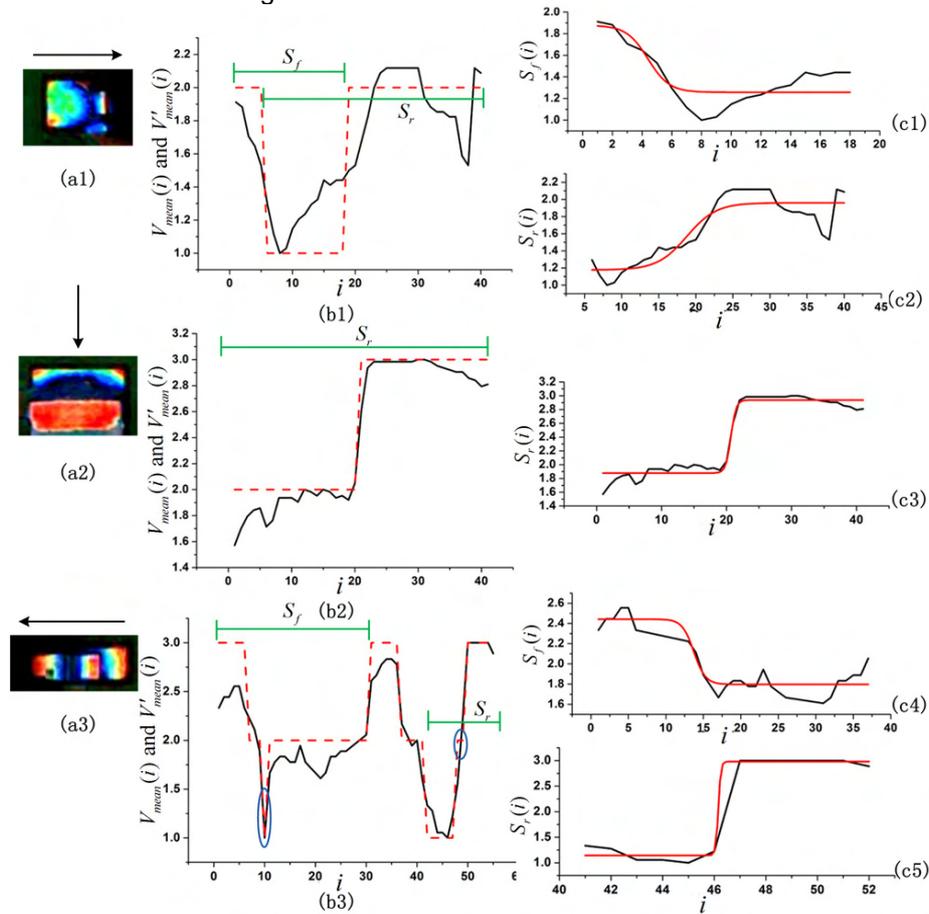


Fig. 5. Color distributions of three typical solder joints. (a1) to (a3) are three typical solder joints; (b1) to (b3) are color distribution curves; (c1) and (c2) are fitting curve of (a1); (c3) is a fitting curve of (a2); (c4) and (c5) are fitting curves of (a3).

Step 4: determine the direction of axis based on the fitted parameters. A variable ϕ_{dir} denotes the direction of axis. It uses 0 to indicate the reverse of

axis, 1 for the forward of axis. The direction of axis is determined according to the following three cases:

In case 1: Only one of S_f and S_r is valid. If S_f is valid, $\phi_{dir} = 0$, otherwise, $\phi_{dir} = 1$. In (b2), only S_r is valid, so, $\phi_{dir} = 1$, the direction of (a2) is along the forward of axis (see black arrow in (a2)).

In case 2: Both S_f and S_r are invalid. ϕ_{dir} is determined by comparison with the length of two data segments. If the length of S_f is larger than that of S_r , then $\phi_{dir} = 0$, otherwise, $\phi_{dir} = 1$.

In case 3: Both S_f and S_r are valid. Over all, there are five kinds of colour sequence. Through observing and analyzing a great deal of color distributions, we set the priority for each sequence. ϕ_{dir} is determined mainly through the priorities. Sequence is evaluated by the fitted parameters A_1 , A_2 and some constraints. Table 1 shows five sequences and their corresponding priority and constraints.

Table 1. Sequence type, priority and constraints

Sequence	Priority	Constraints
{R,B}	3	$A_1 \in [1.5, 2.3) \wedge A_2 \in [2.3, 3.0]$
{R,B,G}	3	$A_1 \in [1, 1.5) \wedge A_2 \in [2.3, 3.0] \wedge (\exists j)(V_{mean}(j) \in [1.5, 2.3)$
{B,G}	2	$A_1 \in [1, 1.5) \wedge A_2 \in [1.5, 2.3)$
{R,G}	1	$A_1 \in [1, 1.5) \wedge A_2 \in [2.3, 3.0] \wedge (\forall j)(V_{mean}(j) \notin [1.5, 2.3)$
others	0	

We evaluate L_f and L_r , the sequence priority of S_f and S_r respectively. If $L_f > L_r$, then $\phi_{dir} = 0$. If $L_f < L_r$, then $\phi_{dir} = 1$. If $L_f = L_r$, ϕ_{dir} is determined by comparison with the length of S_f and S_r . For example, in the case of Fig.5 (b1), $L_f = L_r = 2$, but the length of S_f is less than S_r . Therefore, $\phi_{dir} = 1$, the direction of (a1) is along the forward of axis (see black arrow in (a1)). In the case of Fig.5 (b3), $L_f = 3$ and $L_r = 1$. Therefore, $\phi_{dir} = 0$, the direction of (a3) is along the reverse of axis (see black arrow in (a3)). Fig.6 (a1) and (a2) show the results of determining the direction of solder joints of I_a and I_b , the direction of solder joint is represented by the yellow arrow. Observing the results, we can know that many directions are determined legitimately, and there are still some minor mistakes (error directions are marked by the yellow ellipses). The errors are caused by two reasons: one is the highlight area of a solder joint is incomplete, and it's the main reason; another is the color distribution of some solder joints is so special that we cannot determine its color sequence correctly. However, this is rarely the case.

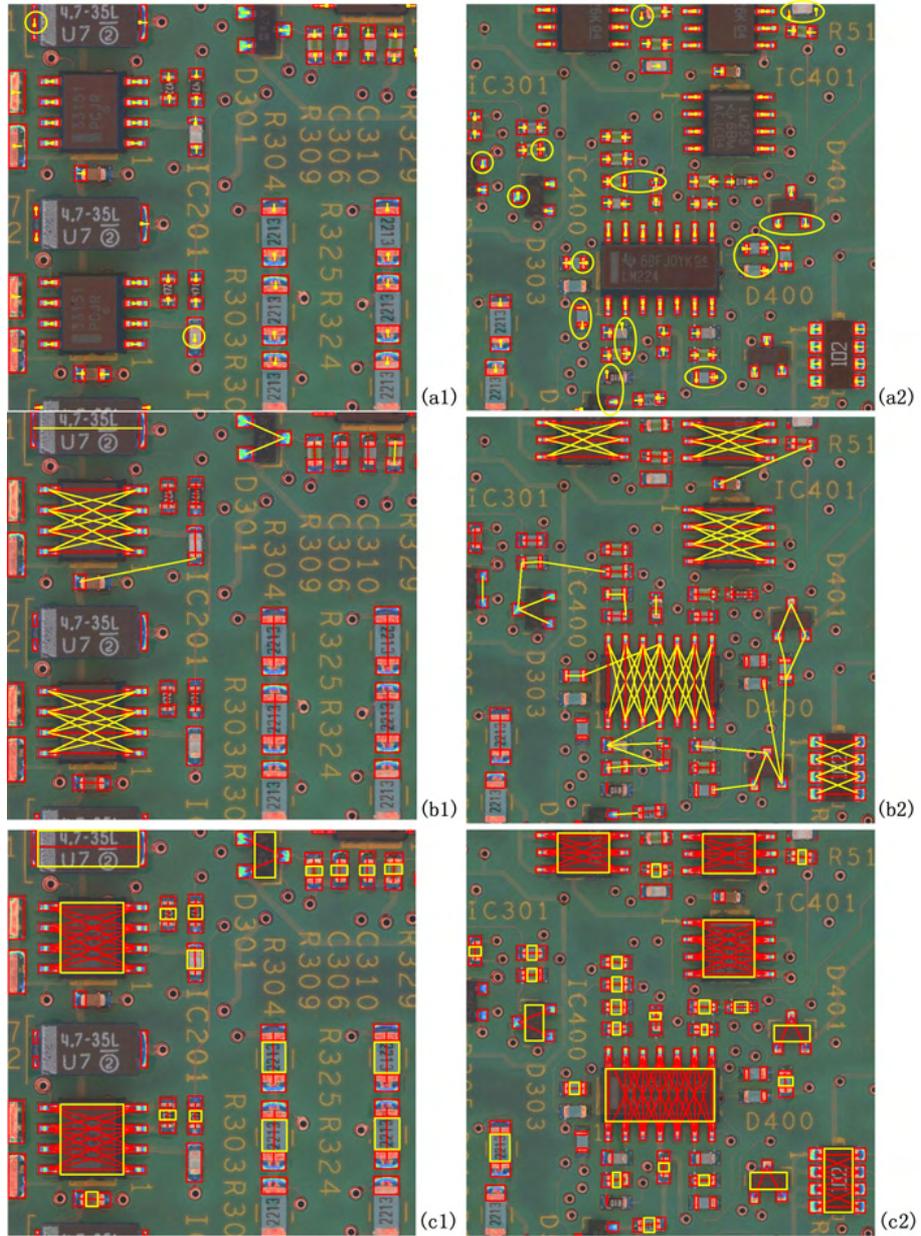


Fig. 6. Clustering and extracting results of I_a and I_b . (a1) and (a2) show the results after determining the direction of solder joints. (b1) and (b2) show the results after clustering solder joints in the first step; (c1) and (c2) show the results after clustering solder joints in the second step and extracting protective coatings.

4.2. Implement details of clustering solder joints and extracting protective coatings

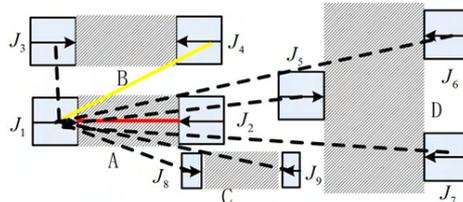


Fig. 7. Some constraints of connections.

The clustering algorithm is divided into two steps: In the first step, we take all solder joints as nodes of a graph, and find all connections between nodes based on the direction of solder joints under some constraints. In the second step, some invalid connections are removed by using the thresholds of PCB background. Finally, the protective coatings are extracted based on the position of clustered solder joints. Some details of clustering algorithm are introduced as follows:

Step 1: Find all connections between solder joints. Suppose that there are two solder joints J_i and J_j , their centers are (u_i, v_i) and (u_j, v_j) respectively, their sizes of highlight area are (w_i, h_i) and (w_j, h_j) respectively. $C(i, j)$ denotes a connection from J_i to J_j . Then $C(i, j)$ is determined by the following constraints: (1) Angle constraint. Compute the angle between J_i and J_j along the direction of J_i , and the angle within $[-45^\circ, 45^\circ]$. (2) Size constraint. Observing many PCB images, we find that all solder joints, which belong to the same component, have the similar size of highlight area. As a result, the relative error $|w_i - w_j| / w_i$ and $|h_i - h_j| / h_i$ ought to be less than ε_1 . In our experiment, we set $\varepsilon_1 = 0.35$. (3) Shortest distance constraint. There are two kinds of the shortest distances. One is along the direction of J_i , we denote it as D_{axis} . Another is satisfy constraint (1) but isn't along the direction of J_i , we denote it as D_{normal} . If all solder joints of a component are distributed symmetrically, then each solder joint of this component has D_{axis} . Otherwise, each solder joint of this component has D_{normal} . For example, in Fig.7, all solder joints of components A, B and C have D_{axis} , all solder joints of component D have D_{normal} . Consequently, the distance between J_i and J_j , $\sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$ must be less than D_{axis} or D_{normal} .

Furthermore, because of the wrong direction of solder joints, not all determined connections are valid. Consequently, we define that if there is a connection $C(i, j)$, and J_j is located along the direction of J_i , and the

where N is the number of pixels, which are within the connection area and also within PCB background. If $N \leq \varepsilon_3$, then $C(i, j)$ is valid. In our experiment, we set $\varepsilon_3 = 0.35$.

Finally, we extract protective coatings. Clustering solder joints according to the valid connections. Each cluster contains all solder joints of one of components. So, the approximate position of a protective coating is evaluated according to the positions of all solder joints in a cluster. Fig.6 (c1) and (c2) show results of clustering solder joints in the second step (see red lines) and extracting protective coatings (see yellow rectangles).

Based on the experimental results, we regard that the introduced algorithm can extract most of the electronic components accurately, except for a few special cases: (1) some components at the sides of image may be incomplete because of the image real-estate constraints. (2) specular area of some solder joints cannot be correctly detected.

5. Conclusion

We present an effective method for extracting components of PCB. The advantages of this method are as follows: First, in order to extract solder joints, we detect all highlight areas first, and then recognize and remove invalid highlight areas which are mainly markings and via-holes. Because the color distribution of markings and via-holes is much simpler than that of solder joints, the proposed approach has higher extraction accuracy in most PCB images. Second, recognizing color object is always a bottleneck, particularly the variation of color within a certain range. We summarize three fundamental features to describe the color distributions of reference object and target object in (u', v') chromaticity coordinate. So, invalid specular areas can be correctly recognized by comparing the distribution features. Finally, the sequence of the color distribution as a new clue has been applied to clustering solder joints. This improves the speed and accuracy of clustering. According to the experimental results, the developed algorithm can effectively recognize most of the chip components. The process of modeling all highlight pixels by GMM is time consuming. Therefore, future research may focus on improving the efficiency of algorithm.

6. Acknowledgement

This project is partly supported by National Science Fund for Creative Research Groups (60821062) 973 Plan (2006CB303105) and NSF of China (60873136)

7. References

1. Bartlet, S.L., Besl, P.J., Cole, C.L., Jain, R., Mukherjee, D., Skifstad, K.D., Automatic solder joint inspection, IEEE Trans. PAMI 10 (1), 31-41 (1988).
2. Bishop, C., Neural Networks for Pattern Recognition. Oxford Uni. Press, (1995).
3. Horng-Hai Loh, Ming-Sing Lu, Printed circuit board inspection using image analysis, *Indust. Automat. and Control: Emerging Technologies*, 673-677 (1995).
4. Chun-Ho Wu, Da-Zhi Wang, A particle swarm optimization approach for components placement inspection on printed circuit boards, *J Intell Manuf*, (2008).
5. WANG Tao, HU Shi-Min, SUN Jia-Guang. Image retrieval based on color-spatial feature[J]. *Journal of Software* ,13(10):2031-2036 (2002).
6. Kim, T.H., Cho, T.H., Moon, Y.S., Park, S.H., Automatic Inspection of Solder Joints Using Layered Illumination, IEEE DOI 9600, pp. 645-648 (1996)
7. Kim, T.H.[Tae-Hyeon], Cho, T.H.[Tai-Hoon], Moon, Y.S.[Young Shik], Park, S.H.[Sung Han], Visual inspection system for the classification of solder joints, *PR*(32), pp. 565-575 (1999).
8. Lee, S.W., Bajcsy, R., Detection of Specularity Using Colour and Multiple Views, *IVC*(10), pp. 643-653 (1992).
9. Tao Lin-Mi, Peng Zhen-Yun, Xu Guang-You. The Features of Skin Color,[J]. *Journal of Software* ,Vol.12, No.7 (2001).
10. Render, R.A., WalkerH.F., Mixture densities, maximum likelihood and the em algorithm. *SIAM Review*, 26(2):195-239, (1984).
11. Shafer,S. A. "Using color to separate reflection components," *Color Res. Appl.*, vol. 10, pp. 210–218, (1985).
12. Lin,S. , Li,Y. , Kang,S. , Tong,X. and Shum,H. "Diffuse-specular separation and depth recovery from image sequences," in *Lecture Notes in Computer Science*, vol. 2352, New York: Springer-Verlag, (2002).
13. Teoh, E.K., Mital, D.P., Lee, B.W. and Wee, L.K., Automated visual inspection of surface mount PCBs, in *Proceedings of IEEE 16th Annual Conference Industrial Electronics Society*, Singapore, pp. 576–580 (1990).
14. Yen, J.C., Chang, F.J. and Chang, S., A new criterion for automatic multilevel thresholding. *IEEE Trans. Image Process. vIP-4*. 370-378, (1995).

Zhou Zeng System Analyst. PhD candidate in Department of Computer Science and Engineering, Shanghai Jiaotong University, China. Reserach interests include computer graphics, image processing and computer vision.

Li Zhuang Ma Full Professor, PhD tutor, and the head of Digital Media Technology and Data Reconstruction Lab. at the Department of Computer Science and Engineering, Shanghai Jiao Tong University, China since 2002. His research interests include computer aided geometric design, computer graphics, scientific data visualization, computer animation, digital media technology, and theory and applications for Computer Graphics, CAD/CAM.

Zuo Yong Zheng PhD candidate in Department of Computer Science and Engineering, Shanghai Jiaotong University, China. Reserach interests include computer graphics and rendering.

Received: April 28, 2009; Accepted: July 04, 2009.

Service-oriented Enterprise Interoperability in Automobile Supply Chain Management

Yong Zhang¹, Shijun Liu¹, Lei Wu¹, and Yuchang Jiao²

¹ School of Computer Science & Technology, Shandong University,
Jinan, P.R.China, 250101
evelyn0330@163.com; {lsj,i_lily}@sdu.edu.cn

² SOA Design Center, China Software Develop Lab, IBM,
Beijing, P.R.China, 100193
jiaoyuchang@gmail.com

Abstract. How to enhance interoperability between stakeholders and improve efficiency of supply chain management is the key issue that needs to be addressed in automobile industry. Existing interoperability solutions are suitable only for large enterprises, however, small and medium-sized enterprises (SMEs) lack cheap, easy to integrate and easy to customize solutions. This paper proposes a methodology that provides a guide on how to establish interoperability between enterprises through a federated approach. An interoperability service platform is designed and delivered in the form of Software as a Service (SaaS). This paper introduces the specifications of the service platform and proposes an interactive framework which is used to establish interoperability between the service platform and other on-premise applications.

Keywords: Enterprise Interoperability, Software as a Service, Supply Chain Management, SOA.

1. Introduction

Today an enterprise's competitiveness is to a large extent determined by its ability to seamlessly interoperate with others. Enterprise Interoperability (EI) has therefore become an important area of research to ensure the competitiveness and growth of enterprises [1].

In modern manufacturing field, such as automobile industry, an entire manufacturing process often cooperated by assembly factory and part suppliers. Agile supply chain management increasingly becomes an effective and important measure to enhance competitive advantage of enterprises that needs the support of agile information system to integrate their supply chain more effectively and quickly [2]. As auto manufacturers inexorably move their sourcing of components and low value-added operations offshore, to lower cost countries, so their supply chains increase in both distance and complexity. Many companies are faced with the challenge of providing an agile

response to customers and yet operating a lean operation across an extended global supply chain. This is a challenge that needs a solution beyond the abilities of simply judgment, the telephone and spreadsheets [3].

Enterprise interoperability issues in the automotive industry are important because the complexity of product, the design process, and the industry magnifies the impact of interoperability problems while obscuring their solutions. However, existing interoperability solutions are suitable only for large enterprises and SMEs lack cheap, easy to integrate and easy to customize solutions. Therefore, SMEs are still far behind in the reform of supply chain management due to their small IT budgets, crude process standards with little visibility data to enable them to share and compare with trading partners.

According to [13] and ISO 14258 [11], three categories of interoperability barriers are identified: conceptual barriers, technological barriers, and organizational barriers. What's more, there are three basic ways to relate systems together to establish interoperations: integrated approach, unified approach, and federated approach.

In this paper, our research work aims at developing a methodology that provides a guide on how to implement an interoperability solution in automobile supply chain management through a service-oriented and federated approach. More precisely the methodology allows establishing interoperability by: (a) constructing a virtual enterprise by identifying and involving various actors and stakeholders in an interoperability service platform; (b) dynamically configuring and composing available interoperability services according to identified requirements; (c) evaluating and improving the interoperability solution in practice.

To enhance interoperability among stakeholders and build agile supply chain for automobile industry, under the guiding of this methodology, an application platform named Supply Business Management (SBM) was designed and delivered in Software as a Service (SaaS) [5] business model, which is cheap, fast, reliable, and without major integration efforts, so they can be invoked by enterprises on the fly in support of their business activities.

In contrast to traditional on-premise software that is deployed and run in a data center at the customer's premise, SaaS software is run at a SaaS hosting provider and can be accessed via the Internet. SaaS offers many advantages for software customers. Instead of software licenses, maintenance and operational costs that occur in the traditional on-premise model, companies consume IT-services in the SaaS model can use the software on demand, just like they would use any other utility such as electricity or water. SaaS user pays only for the usage of the software for example in a pay-per-usage model.

The goal of SBM service platform is to provide a holistic solution enabling the collaborative supply chain management in a flexible and dynamic environment and especially to facilitate SMEs' participation to collaborative supply chain management processes. In SBM service platform, interoperability is considered to be a utility-like capability and delivered in the form of SaaS. This paper introduces the detailed architecture of the platform and proposes an interactive framework which is used to establish interoperability between SBM service and other applications.

2. Related Work

Since the beginning of 2000s, research on enterprise interoperability has been emerging. Most related work is concerned with the elaboration of an enterprise interoperability framework, such as LISI, IDEAS, AIF, EIF, etc. This paper followed the enterprise interoperability framework developed within the frame of INTEROP Network of Excellence (NoE). The purpose of this framework is to define the research domain of enterprise interoperability and help to identify and structure the knowledge of the domain. It has been considered that enterprises systems are not interoperable because there are barriers to interoperability.

In the NoE interoperability framework, there are many research challenges, interoperability service utility is just one of them. The delivery model of interoperability service is a key problem to be considered. On the other hand, Software as a Service (SaaS) is a new delivery model for software, which lowers the cost of development, customization, deployment and operation of a software application to support multiple tenants over the Internet. The SaaS vision focuses on separating the possession and ownership of software from its use. Delivering software's functionality as a set of distributed services that can be configured and bound at delivery time can overcome many current limitations constraining software use, deployment, and evolution.

A well designed SaaS application is generally distinguished by mainly three qualities: scalability, configurability, and multi-tenant efficiency [8] [9]. In [8] four maturity levels for SaaS applications are presented. [15] introduces an electronic contract management application which uses the third level of SaaS maturity model, so the scalability and deployment architecture are not mentioned. However, this paper introduces a SaaS application designed and delivered with the final level of maturity model; both loader balancer and fail over problems are solved.

Variation points have been introduced in [4] [6] [12] and it is one of the key concepts for software product lines to express variability. Variation points, often also called variability points or points of variability, allow the specification of points in a software design that can be customized in order to create new product line members. Similar to software product lines, SaaS applications have variability points where individual customers can customize the application to their needs.

In a SaaS application, the variability points exist in several layers, like UI layer, process layer, data layer, etc. [7] describes how these variability descriptors can be transformed into a WS-BPEL process model that can then be used to guide a customer through the customization of the SaaS application. [4] shows how the service component architecture (SCA) can be extended with variability descriptors and SaaS multi-tenancy patterns to package and deploy multitenant aware configurable composite SaaS applications. While in this paper, in order to realize the configurability, metadata is used to define all the variability points of the application. Meanwhile, JMX is used to manage the metadata so that changed metadata can be hot deployed immediately during runtime.

As far as integration is concerned, in [12] the need for integration of SaaS applications with existing backend systems is motivated and a framework is proposed. [10] uses executable EAI patterns as the solution for EAI problems thus leveraging on the work done in the EAI community for the integration of traditional applications. Our work differs from them as middleware, such as data synchronization toolkit and message engine are designed and developed to address the integration issues in business process layer and data layer.

3. Construction of Virtual Enterprises

The virtual enterprise (V.E.) concept is one of the most important ways to raise the agility and competitiveness of a manufacturing enterprise [13]. Under this concept a master company develops its products by using the manufacturing resources of external partners.

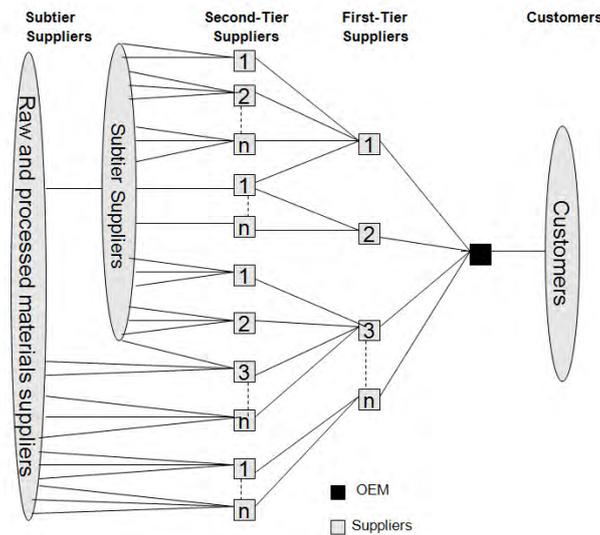


Fig. 1. Typical structure of the supply chain in automotive industry

In the process of designing and manufacturing an automobile, many individuals and organizations exchange product data. The design and manufacturing process involves many divisions within the original equipment manufacturer (OEM), many first-tier suppliers, a number of second-tier and sub-tier suppliers. This exchange of data supports the process of concurrent engineering and design, allowing these organizations to work together to improve the performance and manufacturability of a product and to advance the competitiveness of the industry.

Figure 1 shows a typical structure of the supply chain in automotive industry. As shown in Figure 1, the automotive supply chain consists of four

primary elements: original equipment manufacturers (OEMs), first-tier suppliers, sub-tier suppliers, raw and processed materials suppliers. However, individual companies may operate in several different positions in the supply chain. A company may work for many customers and functions as a first-tier supplier on one project and a sub tier supplier on other projects.

Members of the auto industry generally acknowledge that imperfect interoperability is an important and expensive problem. In order to resolve this problem, the SBM service platform is designed, and the construction of virtual enterprises is a prerequisite to use SBM service. All the stakeholders have to register to be users of the platform, if they want to use the interoperability service utility to support their business activities. After finishing the registration, the OEM users have authorization to create a virtual enterprise. They can choose appropriate suppliers from all the registered users to supply the parts/materials they need for normal production. After the OEM gathering all the suppliers it needs, a virtual enterprise is formally constructed. Then, master of the V.E. could apply for services such as SBM service and APO service. After been approved by the platform, the master could use SBM service with the other members of V.E. to establish interoperability easily and quickly.

It must be addressed that, there exists three different kinds of virtual enterprises in the platform, and the platform can serve all of them well. The first kind of V.E. is the most common and standard one. In this kind of V.E., an OEM is the unique master, and a number of suppliers serve it. The relationship among them is very simple because there is no intersection in the organization. However, in the second kind of V.E., the relationship is more complicated, because intersections appear among V.E.'s boundaries. One supplier may serve several masters. As a result, they will exchange information with different OEMs. The third kind of V.E., is the most complicated one. In the service platform, all the enterprise users could apply for SBM service, so it potentially allows a "double-role" situation. In other words, the master of a V.E., may be a supplier of another V.E.. Just take an engine manufacture enterprise for example, it can be master of a V.E., and get outsourcing parts to produce engines; meanwhile, it may play the role of a supplier to serve other OEMs. According to the analysis of these possible cases, we design and develop the service platform in a flexible way to deal with various possible requirements.

4. Enterprise Interoperability Service Platform

4.1. A Brief Introduction

SBM service platform provides a cooperative environment of supply business between OEMs and their suppliers. With the help of SBM, OEMs can share

production related information easily with their hundreds of suppliers. SBM provides several services to support the daily supply business activities, such as plan service, order service, finance service, quality service, notice service, inventory service etc.

Using the plan service, OEMs could publish their procurement plans to suppliers to guide their production and delivering of materials. There are mainly three kinds of purchase plans, including monthly plans, weekly plans, and daily plans.

Order service may be the most important function of SBM service. OEMs allocate orders to the suppliers according to the pre-defined quota standard. With the help of order service, the decomposed orders are published immediately to corresponding suppliers. Then, the suppliers will print orders online and ship the materials directly to OEMs' warehouses.

In the same way, finance service publishes all the financial information to suppliers, such as general ledger arrearages, quality compensation; inventory service allows suppliers to get the accurate inventory in assembly factory; quality service gives all the quality related data, such as quality inspection, sampling inspection, etc.

However, in order to serve all kinds of OEMs, the SBM service is designed and implemented flexibly, which provides the function of customization. That is to say, users of SBM could customize the service and use it in a DIY way. In order to solve the heterogeneous data integration problems, a data synchronization toolkit, one of the enabling services provided in SBM, is used to synchronize important data from inner systems of an OEM to the database of SBM service. The detailed structure of the synchronization toolkit will be introduced later.

4.2. Architecture Framework

Figure 2 shows the logical architectural framework of SBM service platform designed in SaaS model. This framework provides the capability of hosting multi-tenants using a single application instance supported by the same set of backend servers. In this framework, the presentation and customization modules of the application are deployed in web server, and the Security Service, Metadata Services and Business Services etc, are deployed in application server.

In the framework, business services play an important role in offering the basic functions mentioned above; presentation module deals with the user interface; with the help of metadata services, customization module allows the OEMs to configure the application according to their business needs; security services control the access to OEMs' business data and also ensure its security; shared services such as monitoring service and billing service are used by all the OEMs.

As is shown in Figure 2, a centralized UserInfor database is used to store all the information of users. That's to say, SBM service manages a central user account database that serves all of the application's end users. Each

OEM's administrator is granted permission to create, manage, and delete user accounts in the user account directory. In this way, there's no need to change the OEM's own user infrastructure. As a result, the OEMs could use SBM service more independently. Furthermore, in order to enhance security, besides the traditional username and password, an USB key is required when accessing to the SBM service.

As to the configurability, metadata is used to define all the variation points of SBM. Each of the OEMs has its corresponding metadata files, such as UI.xml, Role.xml, Menu.xml, DBconnection.xml, etc. Moreover, each OEM also has a dedicated metadata file folders to store these metadata files.

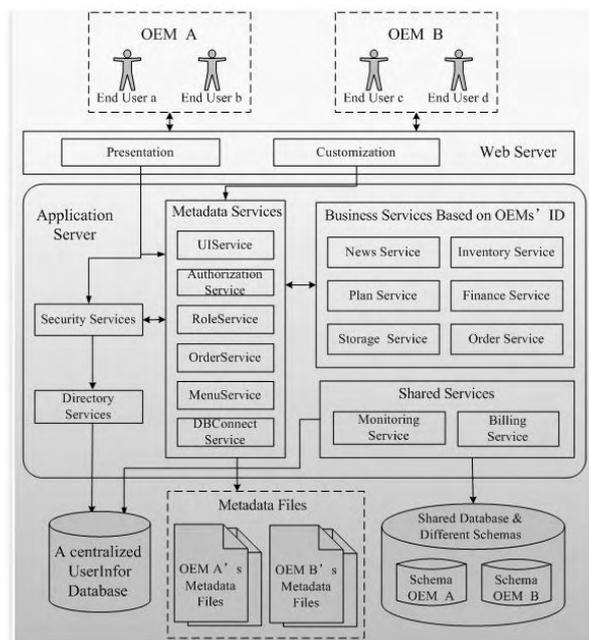


Fig. 2. Architecture framework of SBM service

In fact, metadata services are the most important part in a SaaS application, because: (a) all the variation points of SBM service are customized by OEMs through metadata services; (b) nearly all the business services should use metadata services to parse corresponding metadata files, so as to perform in a desired way; (c) the changes of metadata should be hot deployed and be loaded into the application immediately by metadata services.

Generally speaking, a SaaS application primarily involves three types of data architecture [9]. You will find in Figure 2 that we use the shared database and different schemas to store OEMs' business data. Considering the hardware cost and the security of data, we finally choose this kind of data architecture. With the support of SBM service, assembly factory could

enhance and improve interoperability with its suppliers with little IT investment and operational cost, instead of developing a new solution from scratch.

4.3. JMX-based Metadata Management

As is mentioned above, all the variation points of SBM service are defined with metadata and stored in the form of XML. As a result, each OEM has a corresponding file folder to store its metadata files.

In order to realize the configurability of SBM, Metadata Services are used to parse these metadata files and load the customized information into the application. At present, there are mainly two ways to load metadata files: (a) load the metadata files when the application is initialized; (b) load the metadata files when they're required. Because almost all the functions of SBM service could be customized by OEMs, if the metadata files are loaded when they're required, the performance of SBM will be severely affected. Therefore, we load all the metadata files at startup, and all needed information is stored in memory, and the performance will be higher.

In practice, OEMs may often change some of the metadata files during runtime. Because this application is critical to the business, it needs to be configurable without causing a shutdown of operations. That's to say, you cannot restart the application to reload the changed metadata files, because multi-tenants are using the same instance. OEM A's customization of metadata files must be transparent to OEM B. A general way to solve this problem is to reread and reload the changed metadata files after they're modified and saved. The advantage of this way is that, you need not restart the application. However, the disadvantage is that, the changes of metadata cannot work immediately until the changed metadata files are reread and reloaded. What's more, even if a single line of the metadata file is changed, the whole metadata file must be reloaded. As a result, the users will get a bad performance.

In order to thoroughly solve this problem, we used the JMX[14] to realize the hot deployment of metadata. With the help of JMX, only the changed metadata will be reloaded immediately to the memory, and won't have to reload the whole metadata files.

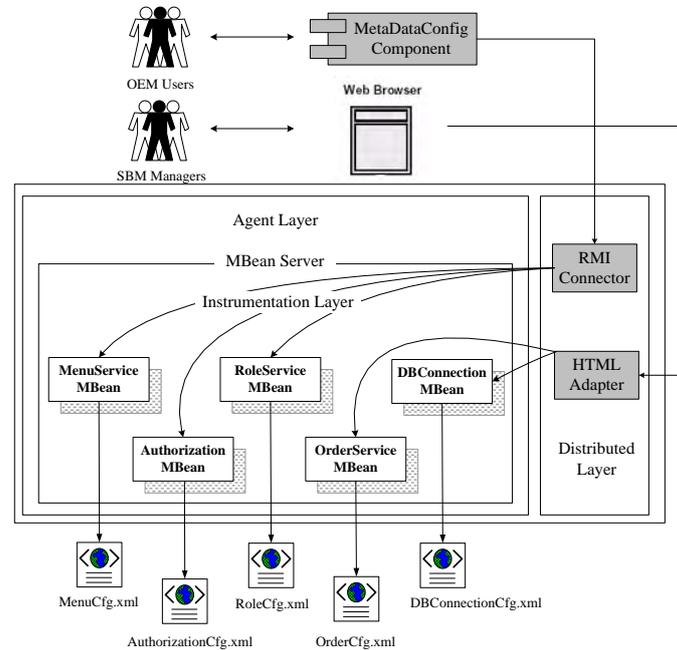


Fig. 3. JMX based metadata management

As is shown in Figure 3, the JMX architecture has three layers: instrumentation layer, agent layer, and distributed layer. JMX uses Java classes called managed beans (MBeans for short) to expose predefined portions of one application. In SBM service, every metadata file has a dedicated MBean defined and used to manage it. For example, the MenuServiceMBean is used to manage MenuCfg.xml. Furthermore, MenuServiceMBean can be defined to control all the elements and attributes of MenuCfg.xml, such as 'menu_id', 'menu_name', etc. Consequently, only the changed elements or attributes in MenuCfg.xml will be hot deployed directly into memory, the whole file won't be reloaded.

In JMX, all the MBeans have to be registered to the MBean Server in agent layer. Management tools access these MBeans by interacting with JMX agents that make the MBeans available to any number of protocols and technologies such as SNMP, Java RMI, and HTTP. It is clear that in Figure 3, some of the MBeans are used by the OEMs users to customize metadata files. These MBeans are connected with RMI connector by Metadata Configure Component. On the other hand, the rest of MBeans are used by the SBM managers to manage the runtime behavior, for instance, dynamically modifying the size of database connection pool. These kinds of MBeans are connected by HTML adapter, and can be managed directly through a web browser.

5. Scalable Deployment Architecture

A well designed SaaS application can be scalable to an arbitrarily large number of customers, and the number of servers and instances on the back end can be increased or decreased as necessary to match demand, without requiring additional re-architecting of the application, and all the changes can be rolled out to thousands of tenants as easily as a single tenant. Scalability is one of the key problems that need to be addressed. Generally speaking, the scalability of a SaaS application is often discussed in application layer and database layer.

5.1. Scalability in Application Layer

SBM service is designed to be scalable, fast, reliable and able to handle fail-over behavior. However, with an increased number of end users, the performance of SBM degrades and it is necessary to distribute client requests to different servers in order to perform parallel processing. Client requests towards web servers are distributed using a common front-end load balancer and high availability is achieved using a set of clustered servers with the same set of files and similar configuration.

In the construction of server cluster, the use of session must be solved. Generally speaking, two ways are widely used to deal with session: session replication and session sticky. Session replication is time-consuming and may block the network. On the other hand, session sticky is simple to implement, but it cannot solve the fail over problem.

Shared nothing architecture [16] (SN) is a distributed computing architecture in which each node is independent and self-sufficient, and there is no single point of contention across the system.

As is indicated in Figure 4, according to SN architecture, the SBM service is designed to run in a stateless fashion, all the session data is stored in a central session sever which is accessible to any application instance. In this way, each transaction can be handled by one instance as well as any other; all the instances are equal with each other. So if one of the instances is down, the requests of users will be forward to another instance and the process is transparent to users. Both the load balance and the fail over requirements are addressed.

5.2. Scalability in Data Architecture

Data architecture is an area in which the optimal degree of isolation for a SaaS application can vary significantly depending on technical and business considerations. [9] introduces three approaches to managing multi-tenant data: separate databases; shared database, separate schemas; and shared database, shared schema.

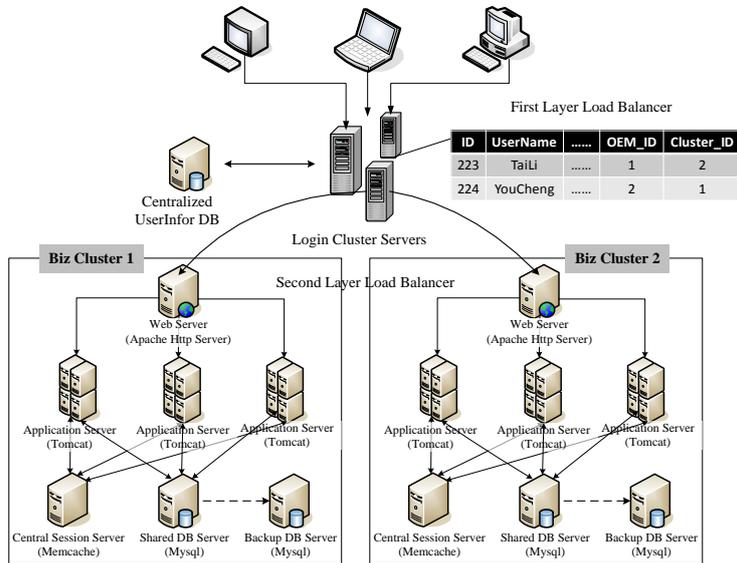


Fig.4. Business cluster based deployment architecture

Considering hardware and maintenance costs and data isolation requirements, the data architecture of SBM service follows the approach of shared database and separate schemas. The separate-schema approach is relatively easy to implement, and OEMs can extend the data model easily. This approach offers a moderate degree of logical data isolation for security-conscious OEMs, and can typically accommodate more OEMs at a lower cost.

As databases serve more users concurrently and grow in size, the amount of time it takes to perform operations such as querying and searching increases significantly. Therefore, a shared database should be scaled to accommodate more tenants or heavier usage when it can no longer meet baseline performance metrics. However, the separate-schema approach is well-suited to the tenant-based horizontal partitioning pattern because each OEM has its own set of data, so the managers can easily target individual tenant data and move it to another server. You can also find in Figure 4 that in order to ensure the high availability in data layer, we used a real time backup DB server to backup the production server.

5.3. Business Cluster-based Deployment Architecture

In general, in order to improve the scalability of a SaaS application, both the application server layer and the database layer should be scaled out. Assume that, if there are M application servers and N database servers in SBM service, the connections among them will be M multiple N. On the other hand, the database connections are very precious resources, when the operations

such as querying and searching increases significantly, the processing ability of the application will be in danger.

So we propose a *business cluster* based deployment architecture to solve this problem, as is indicated in Figure 4. In a business cluster, an application server cluster is bound with a concrete database server, and all the owners of the data stored in the database are bound with the application server cluster. As a result, an application server cluster is dedicated with a number of end users and their shared database server.

As is shown in Figure 4, requests from all the users are processed by the Login Server Cluster, which contains a centralized authentication system. In the UserInfor DB, each OEM is bound with a dedicated business cluster. As a result, all the requests from end users will be first dispatched to corresponding business clusters, so the Login Server Cluster works as the first layer loader balancer. However, within a business cluster, the requests will be dispatched by the web server which works as the second layer loader balancer.

As we can see from Figure 4, the SBM service can scale out almost infinitely simply by adding nodes in the form of inexpensive computers, since there is no single bottleneck to slow the system down. When the number of users is increasing rapidly, the deployment architecture can also meet the requirement of scalability. The business cluster based deployment architecture has been proved to be flexible and efficient.

6. Interoperability Challenges and Solutions

6.1. Integration Requirements

Most of the time, SaaS applications are only one small part of the overall IT-landscape of an enterprise. As far as it goes, most medium and large OEMs have certain applications already deployed on their premises. There's no question SaaS integration problems do exist: all the OEMs expect that SBM service can be integrated with their legacy applications. However, these kinds of integration requirements must be resolved; otherwise, the customers will leave and find another service provider. In the operation of our application, we also met the problem of integration. The integration requirements mainly exit in the process layer and data layer.

A business process supported by SBM service usually can trigger business process supported by another on-premise application. For example, an order process from SBM service would trigger an order fulfillment process managed by an ERP application. Therefore process integration can automate the end to end business process transaction spanned across SBM service and on-premise applications.

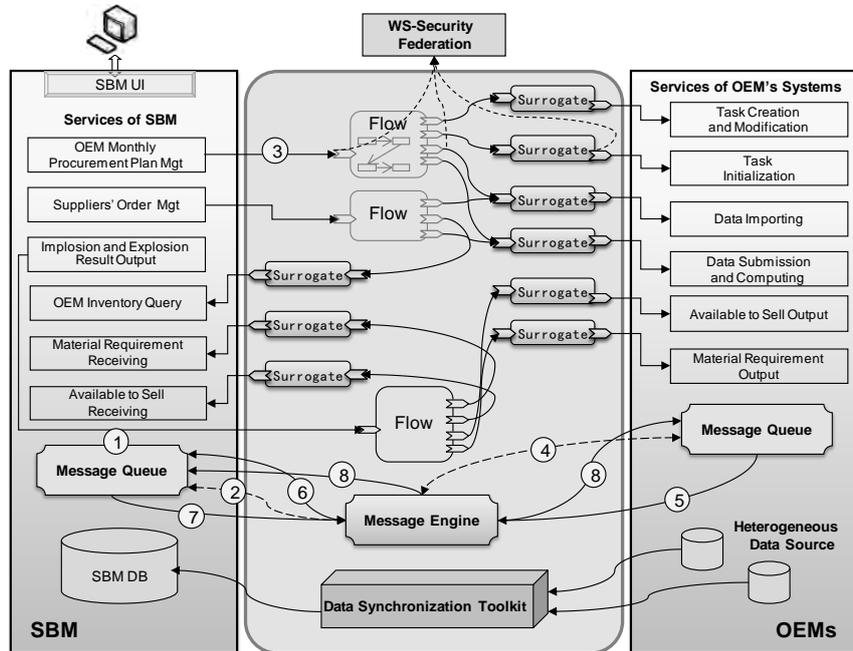


Fig.5. Integration roadmap between SBM service and on-premise application

There are two types of data in SBM service: master data and transactional data. These data should be synchronized from SBM services to OEM's on-premise applications and vice versa.

Figure 5 illustrates the integration roadmap we adopt to meet these integration requirements. In order to address the issues of data integration, we designed and developed a data synchronization toolkit to synchronize the important data between the SBM service and the heterogeneous data sources within OEMs. Furthermore, service surrogate extended from SCA is used to connect SBM service with the of inner systems, and in order to ensure the interoperability between them, a tool named Message Engine is used to provide coordination functions.

6.2. Data Synchronization Toolkit

As is shown in Figure 6, a data synchronization toolkit is used to synchronize the important data between the SBM service and the heterogeneous data sources. The synchronization toolkit has two parts: the client side is deployed within OEMs, and the server side is deployed in SBM service. The client side is used to extract the needed data from heterogeneous data sources and send them to the server side. On the other hand, the server side is used to receive and load the data into SBM service.

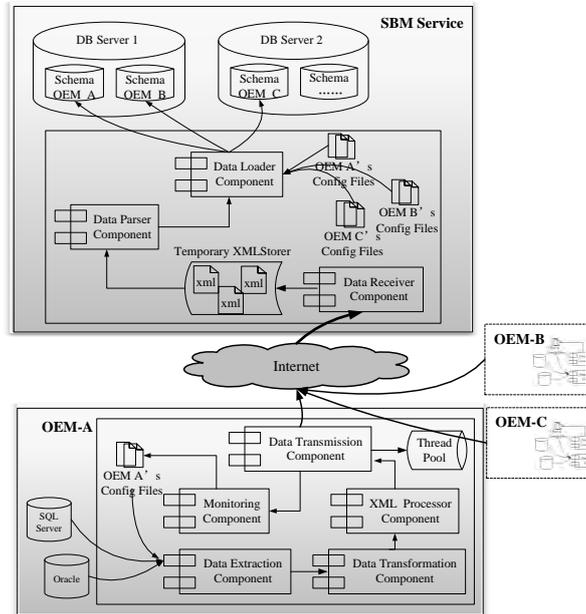


Fig.6. Data synchronization toolkit

It must be pointed out that, the data synchronization toolkit also works in a SaaS way. Because OEM's data may be stored in different databases and in different structures, the synchronization requirements must be different. So the data synchronization toolkit is designed and delivered with configurability: both the client and the server side provide a set of configure files which can be customized by OEMs according to their individual situation.

The client side deployed within OEMs will first extract the needed data from heterogeneous data sources according to the configure files. Then, all the data will be transformed into XML files. As a result, the difference among heterogeneous data sources is shielded. After that, a dedicated XML processor component is used to split and compress these XML files. Finally, the processed XML files will be transferred through internet.

On the server side, the received XML files will first be stored in a temporary file folder and then be parsed and loaded into dedicated schemas according to the configure files. The whole process will be monitored, in order to deal with some unpredictable situations.

6.3. Service Connection

In order to ensure the interoperability in process layer, several technological issues have to be addressed: (a) how to connect the web services exposed by SBM service and on-premise applications to execute business process and

get the needed data; (b) how to coordinate these services and make them cooperate correctly.

Generally speaking, there are various services in a heterogeneous environment; web service is just one of them. How to identify them and make them work together is an issue need to be addressed. This paper provides solutions including service authentication, service authorization and service connection according to different business requirements.

Service surrogate extended from SCA [18] provides a programming model which is used to construct applications and solutions based on SOA. SCA (Service Component Architecture) is a model that aims to encompass a wide range of technologies for service components and for the access methods used to connect them.

As to access methods, SCA compositions support various communication and service access technologies such as web service, messaging system and RPC (Remote Procedure Call). The original definition of SCA Component defines the implementation, exposure and invoking mechanism of services. However, it cannot fulfill our requirements for interoperability, such as service authorization and service authentication.

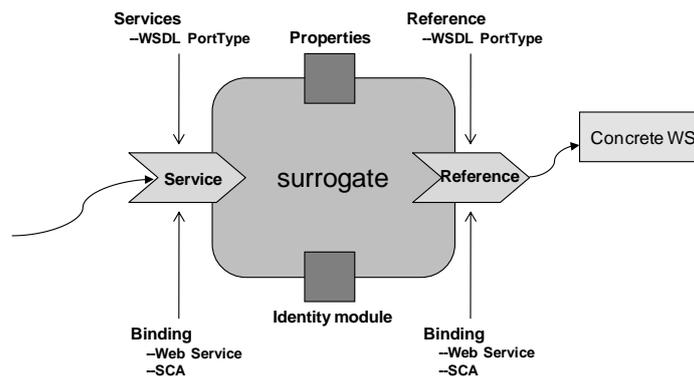


Fig.7. Definition of surrogate based on SCA component

SCA provides service flow definition mechanism to carry out complicated business processes automatically. It is the same as the extended SCA component. SCA describes the content and linkage of an application called composites. Composites can contain components, services, references, property declarations, plus the wiring that describes the connections among these elements. Composites can group and link components built from different implementation technologies, allowing appropriate technologies to be used for each business task [9].

This paper extends SCA Component by adding the Identity Module to monitor and process SOAP messages based on WS-Security. As is revealed in Figure 7, the extended SCA Component is named "surrogate". In general, SBM service and on-premise applications manage their users separately. So the detailed solution of service connection is:

- 1) Encapsulating the related web services with SCA Component.
- 2) Process the authentication information in the identity module.

A mapping table for users should be maintained in SBM service. The table stores the mapping relationship of SBM users and on-premise application users. The role and authorization will not be changed and still be managed separately. When an authorized user, such as Bob logs in SBM service and invokes services provided by on-premise application, the authentication process will be activated. Firstly, Identity Module will insert Bob's identity information into SOAP head; then the gateway will capture the message and find the corresponding user of Bob in the mapping table. If succeeds, the new user_id certified will be inserted into SOAP head to invoke the services exposed by on-premise application. The return process is carried out in the same way.

6.4. Service Coordination

Service coordination is often used to support a number of applications, including those that need to reach consistent agreement on the outcome of distributed transactions. In order to ensure the interoperability between SBM service and on-premise application, this paper proposes a Message Notification based service coordination technology, which is driven by business rules. We design and implement a tool named Message Engine in dependence on Apache Kandula2 [17] project, which provides an open-source implementation of WS-Coordination, WS-Atomic Transaction, and WS-Business Activity.

The Message Engine not only enables SBM service to create a context needed to propagate an activity to on-premise application, but also enables the coordination of transactions among them.

As is shown in Figure 5, with the help of the Message Engine, the whole process of service coordination based on Message Notification is carried out by the following steps:

- 1) Configuring the Message Queue of SBM service and on-premise application;
- 2) Registering the Message Queue of SBM service to Message Engine;
- 3) Users log in SBM service and then activate a business process;
- 4) Web services involved in the process are invoked and Message Queue of on-premise application is registered to Message Engine;
- 5) Message Queue in on-premise application invokes the completed() function automatically after the execution;
- 6) Message Engine invokes the complete() function of SBM to notify the end of the process;
- 7) On receiving the notification, the whole process ends;
- 8) Notifying on-premise application the whole process ends.

With the help of these technologies, interoperability between SBM and on-premise application can be established to fulfill the requirement of making them work together.

7. Conclusions and Future Work

This paper proposed a methodology that provides a guide on how to establish interoperability between enterprises through a federated approach. Under the guide of this methodology, the paper designed and realized an interoperability service platform to enhance the interoperability and efficiency of the supply chain management in automobile industry. The architecture framework of SBM is introduced, and metadata is used to define all the variability points. In this paper, JMX is used to manage all the metadata files. In addition, shared nothing architecture is used to design the business cluster based deployment architecture. Data synchronization toolkit and message engine are implemented to address the integration issues in business process layer and data layer.

As far as it goes, the SBM service has been used by more than 1,200 suppliers, and the number is increasing. With the help of SBM service, interoperability between OEMs and suppliers is constructed and enhanced. On average, nearly 10,000 orders are transported by the data synchronization toolkit every day. However, there are also several problems need to be addressed in the future. For example, the data synchronization toolkit needs to be improved, because it'll be influenced easily by the condition of network. Meanwhile, we should do more research in the security aspect to ensure the security of tenant's data. Besides, more work will be done in the research of SLAs and QoS of services.

8. Acknowledgement

The authors would like to acknowledge the support provided for the project by the National Natural Science Foundation of China (60703027), the National High Technology Research and Development Program of China (2006AA01A113, 2009AA043506), the Science & Technology Development Projects of Shandong Province (2007GG10001001), the Natural Science Foundation of Shandong Province (ZR2009GM028), and the China's Post-Doctoral Science Fund (20090451313).

9. References

1. Yannis Charalabidis, George Gionis, Karl Moritz Hermann, Cristina Martinez; "Enterprise Interoperability Research Roadmap";Feb.2008
2. G arokh, M. J.; Ghahremanloo, Hoda; Karami, Mahnaz, Agility in Auto Dealers SCM, Service Operations and Logistics, and Informatics, 2007. SOLI 2007. IEEE International Conference on Volume, Issue , 27-29 Aug. 2007 Page(s):1 – 6
3. Ross, A. Creating agile supply chains, Manufacturing Engineer, Dec. 2003, Volume: 82, Issue: 6, pp: 18-21

Yong Zhang, Shijun Liu, Lei Wu, and Yuchang Jiao

4. Ralph Mietzner, Frank Leymann. Defining Composite Configurable SaaS Application Packages Using SCA, Variability Descriptors and Multi-Tenancy Patterns. The Third International Conference on Internet and Web Applications and Services.
5. Abhijit Dubey and Dilip Wagle, "Delivering software as a service", The McKinsey Quarterly, Web exclusive, May 2007
6. J. Bosch. Design and use of software architectures: adopting and evolving a product-line approach. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 2000.
7. R. Mietzner and F. Leymann. Generation of BPEL Customization Processes for SaaS Applications from Variability Descriptors. In IEEE International Conference on Services Computing, 2008.
8. F.Chong and G. Carraro. Building Distributed applications: Architecture Strategies for Catching the Long Tail. MSDN architecture center.
9. F.Chong and G. Carraro. Building Distributed applications: Multi-Tenant Data Architecture. MSDN architecture center.
10. Thorsten Scheibler, Ralph Mietzner. EAI as a Service - Combining the Power of Executable EAI Patterns and SaaS. 12th International IEEE Enterprise Distributed Object Computing Conference
11. ISO 14258, Industrial Automation Systems—Concepts and Rules for Enterprise Models, ISO TC184/SC5/WG1, April 14, 1999.
12. M.Jaring and J. Bosch. Architecting product diversification- formalizing variability dependencies in software product family engineering. In QSIC '04: Proceedings of the Quality Software, Fourth International Conference, pages 154–161, Washington, DC, USA, 2004. IEEE Computer Society.
13. D. Wang, K.L. Yung, W.H. Ip, "A heuristic genetic algorithm for subcontractor selection in a global manufacturing environment", IEEE Transactions on Systems, Man and Cybernetics—Part C: Applications and Reviews, 2001, 31 (2), pp.189–198.
14. BEN G. SULLINS. JMX in Action. Manning Publications, ISBN 1-930110-56-1, 2003.
15. Thomas Kwok, Thao Nguyen. A Software as a Service with Multi-tenancy Support for an Electronic Contract Management Application. In IEEE International Conference on Services Computing, 2008.
16. http://en.wikipedia.org/wiki/Shared_nothing_architecture
17. <http://ws.apache.org/kandula/2/index.html>
18. SCA Service Component Architecture – Assembly Model Specification. <http://www.osoa.org/display/Main/Service+Component+Architecture+Home>

Yong Zhang has received his BS and MS degree in software engineering from software college, Shandong University, China. Currently, he is enrolled at the school of computer science and technology, also in Shandong University, as a PhD candidate working on services computing, grid computing. His current research interests include SOA, Web Service, grid computing and services oriented computing, enterprise interoperability.

Shijun Liu is an associate professor of the School of computer science and technology at Shandong University, China. He earned his BS degrees in Oceanography from Ocean University of China, and MS and PhD degree in computer science from Shandong University. His teaching and research interests are focused on Service Computing, Networked Manufacturing, Computer Integrated Manufacturing System and Manufacturing Grid. He has published 2 books and over 50 papers in journals and conferences and obtained many academic awards including the Nation Scientific and Technological Progress Award in 2002.

Lei Wu has received her Ph.D. degree in computer software and theory from Shandong University, China. Now, she is a lecturer at the School of Computer Science and Technology at Shandong University. Her research interests are focused on networked manufacturing; computer integrated manufacturing systems and manufacturing grids.

Yuchang Jiao has received his master degree in software and theory from Shandong University in 2008. He is currently working at IBM SOA Design Center, Beijing, China. His current research interests include Service Oriented Architecture, services computing and virtualization technologies.

Received: April 14, 2009; Accepted: September 26, 2009.

Modeling and Simulation of a Spherical Mobile Robot

SANG Shengju^{1,2}, ZHAO Jichao², WU Hao¹, CHEN Shoujun¹, and AN Qi¹

¹ School of Mechanical and Power Engineering, East China University of Science and Technology, Shanghai, P.R. China, 200237;
sang1108@163.com, wh@mail.ecust.edu.cn, csj19850112@yahoo.com.cn, anqi@ecust.edu.cn

² School of Information Science and Technology, Taishan College, Shandong Taian, P.R. China, 271021
{sang1108, zjc}@163.com

Abstract. Spherical mobile robot (SMR) has been studied analytically and experimentally in this paper, a novel design with an internal propulsion mechanism and mathematical models of the robot's dynamics and kinematics are introduced. A 3D model of robot is built by SOLIDWORKS and then exported to ADAMS2007 for simulation. The results of simulation by combining MATLAB / SIMULINK with ADAMS are presented. It is shown experimentally that the behavior of actual model consist well with the prediction of simulation.

Keywords: spherical mobile robot (SMR), simulation, ADAMS, virtual prototype, kinematics and dynamic, mathematical model.

1. Introduction

Spherical mobile robot (SMR) is a new type of robot [1]-[4], which has a ball-shaped outer shell to accommodate all its mechanism, control devices and energy sources in it. Spherical robot is characterized as simple, compact, well-sealed structure and agile motion and it has attracted the interest of many researchers. The spherical structure offers extraordinary motion properties in cases where turning over or falling down will bring risks during motion. A spherical mobile robot which has full capability to recover from collisions with obstacles can be used to survey unstructured hostile industrial environment or explore other planets [9]. In addition, a spherical mobile robot can quickly move as a wheeled robot and also can easily detour obstacles as a legged robot.

This paper presents a novel robot that can achieve many kinds of unique motion, such as all-direction driving on rough ground, without great loss of stability. The structure of the robot presented in this paper is one of spherical mobile robots, which can be seen in Fig. 1.

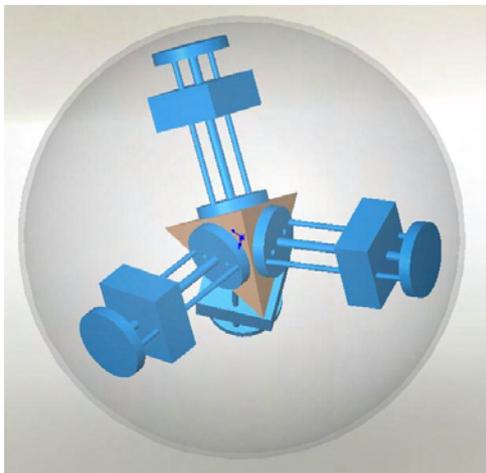


Fig. 1. Prototype of the spherical robot

2. Related Works

The spherical mobile robots have been studied by using a variety of mechanisms [1-8]. In other words, several omnidirectional platforms have been known to be realized by developing a specialized wheel or mobile mechanism. Mobile robots of spherical shape have been described by only a few authors. The first spherical robot was developed by Halme et al. [1]. They proposed a spherical robot with a single wheel resting on the bottom of a sphere. Bicchi et al. [3] developed a spherical vehicle consisting of a hollow sphere with a small car resting on the bottom. Bhattacharya et al. [10] proposed a driving mechanism that is a set of two mutually perpendicular rotors attached to the inside of the sphere. Ferriere et al. [7] developed a universal wheel to actuate a spherical ball to move the system, and in their mechanism, the actuation system is out of the sphere. It is, however, obvious that the autonomous spherical rolling robot, described in this paper, glory, differs from the previous designs. The driving mechanism is a set of four spokes distributing weights radially along them inside the sphere, in contrast to a wheel resting at the bottom of the sphere or, two mutually perpendicular rotors attached to the inside of the sphere, in previous works. Mukherjee et al. [11] studied the feasibility of moving a spherical robot to an arbitrary point and orientation. They had presented two strategies for reconfiguration. Glory is the only implemented spherical robot which can traverse omni-directionally [4]. The review papers discussing constructional details of available spherical mobile robots are [10, 11] and a concise review on the existing path planning algorithms and feedback algorithms for the system can be obtained from [12,

13]. In this paper, we present a systematic study of a spherical mobile robot, designed, fabricated and analyzed in our laboratory.

The organization of this paper is as follows. Related work is reviewed in section 2. Section 3 describes the construction and design in detail. Section 4 describes the mathematical modeling of the system using quaternion. In Section 5, the trajectory of the robot is discussed in the quaternion space. The simulation and experimental setup along with the discussion on the experimental results is reported in section 6. Section 7 provides concluding remarks.

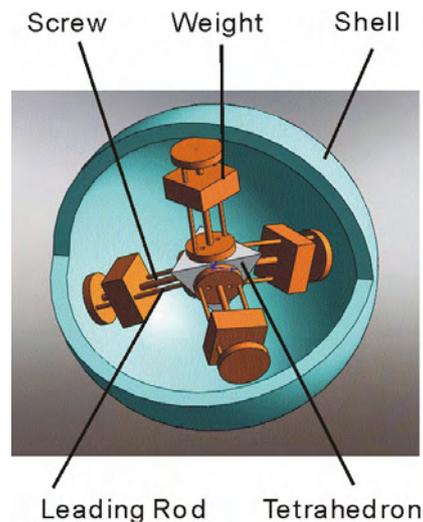


Fig. 2. Structure of the spherical mobile robot

3. Design

The spherical mobile robot presented in this paper has an external spherical shape. It is composed of a spherical shell, a Propulsion Mechanism and a stable tetrahedron. Fig.1 shows its 3D model and Fig.2 shows the structure of the spherical mobile robot.

The spherical shell is made up of acrylic material having 4 mm thickness. The inner radius of the robot is 360mm. The transparent acrylic spherical shell enables researchers to monitor the state of internal mechanism while in motion. The spherical robots work on the principle of change in the center of gravity. A crucial aspect of the design is to place the internal components to make the center of mass of the robot exactly at the geometric center of the sphere. This is very important so that the robot will not tip over on its own. The easiest way to achieve this is to place all the internal components symmetrically [18, 19]. It is absolutely vital that there be no relative motion

between the two hemispheres in motion. This can be achieved screwing a connecting rod along the axis of the sphere.

The propulsion mechanism consists of four power screwed spokes, connected in 109.47° inside a tetrahedral shape [4], There are four heavy objects (heavy shortly or weight), placed through spokes, which are elevated upward and downward using four stepper motors, with 200 steps per revolution, connected directly to the spokes. The motor that is driving the heavy is also fixed on the hollow shaft and tied to the heavy by a link.

4. Mathematical Modeling

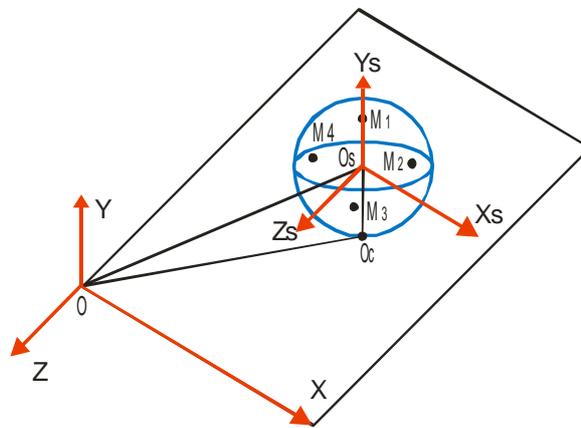


Fig. 3. Coordinates setup of the spherical mobile robot

This section describes the development of an analytical model of the spherical rolling robot using quaternion. Considering a spherical robot, rolls on a horizontal plane as shown in Fig.3, an inertial coordinate frame is attached to the ground and denoted as XYZ with its origin at the point O . The body coordinate axes $X_sY_sZ_s$, parallel to XYZ , are attached to the sphere and have their origin at the center of the sphere O_s . The set of generalized coordinates describing the sphere consists of 1) Coordinates of the contact point O_c on the plane, and 2) any set of variables describing the orientation of the sphere [14]-[17].

We use Euler parameters (instead of Euler angles) which is a set of 4 parameters to describe the orientation of the sphere. Euler parameters have the advantage of being a nonsingular two to one mapping with the rotation. In addition, Euler parameters form a unit quaternion and can be manipulated using quaternion algebra [20, 21].

Because the spherical robot cannot move in Y direction, five variables $(X, Z, \alpha, \beta, \gamma)$ are enough to describe its space state. x and z are the

coordinates of the contact point Oc between the sphere and the plane and (α, β, γ) are the generalized Euler angles used to describe the sphere orientation. We have adopted a Newton formulation for propulsion mechanism since the acceleration of the sphere appears explicitly in the equation. These equations have the form:

$$\sum \bar{M} = (\sum [I]) \cdot \bar{\omega} + \sum \bar{\rho} \times m \bar{\rho} \quad (1)$$

Where $(\sum [I])$ is the moment due to the weights of the individual unbalanced masses, I is the moment of inertia of the robot and $\bar{\omega}$ is the angular acceleration of the sphere. The vector $\bar{\rho}$ represents the position vector of i th weight with respect to the relative spoke. The vectors represented by $\bar{\rho}$ are equivalent to the absolute accelerations of individual weights.

In our implementation, due to discrete nature of stepper motors, we can suppose, for each step of simulation, each heavy is temporarily fixed in its position related to the geometry. Thus we can neglect the second term on the right hand of the equation in a good rate of approximation. It helps calculating angular acceleration. A static analysis only considers the moments due to the weights of the unbalanced masses and has the following form:

$$\sum \bar{M} = (\sum [I]) \cdot \bar{\omega} \quad (2)$$

Let i, j, k be the unit vectors of the body frame and ω_s be the angular velocity of the sphere given by

$$\omega_s = (\dot{\alpha} c \beta c \gamma + \dot{\beta} s \gamma) i + (-\dot{\alpha} c \beta s \gamma + \dot{\beta} c \gamma) j + (\dot{\alpha} s \beta + \dot{\gamma}) k \quad (3)$$

Where $c \beta$ and $s \beta$ are short for $\cos \beta$ and $\sin \beta$ respectively.

Let ξ, η, ζ be the unit vectors of inertial coordinate frame, and the projection of the angular velocity vector on the body axes can be derived and expressed in the following form.

$$\begin{aligned} \omega_s &= \omega_\xi + \omega_\eta + \omega_\zeta \\ &= (\dot{\alpha} + \dot{\gamma} s \beta) \xi + (\dot{\beta} c \alpha - \dot{\gamma} s \alpha c \beta) \eta + (\dot{\beta} s \alpha + \dot{\gamma} c \alpha c \beta) \zeta \end{aligned} \quad (4)$$

We assume that the robot rolls without slipping. Hence, the velocity of the contact point Oc, with respect to the inertial coordinates is zero, i.e. $v_{Oc} = 0$. The constraint equations reduce to

$$v_r^\xi = \dot{x} + \omega_\zeta r = \dot{x} + (\dot{\beta} s \alpha + \dot{\gamma} c \alpha c \beta) r \quad (5)$$

$$v_r^\zeta = \dot{z} - \omega_\xi r = \dot{z} - (\dot{\alpha} + \dot{\gamma} \sin \beta) r \quad (6)$$

Where r is distance vector.

Due to the symmetric design of the ball, the gravity force acts vertically through the center of the spherical robot and the contact point Oc. Both the ground reaction force and frictional force act through the contact point. Hence, the sum of the external moment at the contact point Oc is zero. And the angular momentum of the robot at Oc is a conservative quantity. Hence, weights are the only bodies that cause moment and define the movement direction with their inter-relation. Therefore, the resultant moment vector is as following.

$$\bar{M} = mg \begin{bmatrix} z_1 + z_2 + z_3 + z_4 \\ 0 \\ -x_1 - x_2 - x_3 - x_4 \end{bmatrix} \quad (7)$$

Where x_j and z_j are the position of the j^{th} weight ($1 < j < 4$).

5. Trajectory Planning

Because of omnidirectional property of the robot, direct path is proposed and the traveling path is very close to the shortest path to the target. For calculating the moment of inertia we use three 3X3 matrixes as below:

$$I_j^w = \begin{bmatrix} I_{jxx}^w & I_{jxy}^w & I_{jxz}^w \\ I_{jyx}^w & I_{jyy}^w & I_{jyz}^w \\ I_{jzx}^w & I_{jzy}^w & I_{jzz}^w \end{bmatrix}, I_j^c = \begin{bmatrix} I_{jxx}^c & I_{jxy}^c & I_{jxz}^c \\ I_{jyx}^c & I_{jyy}^c & I_{jyz}^c \\ I_{jzx}^c & I_{jzy}^c & I_{jzz}^c \end{bmatrix}, I^s = \begin{bmatrix} I_x^s & 0 & 0 \\ 0 & I_y^s & I_{yz}^s \\ 0 & 0 & I_z^s \end{bmatrix} \quad (8)$$

Where I_j^w, I_j^c and I^s are the moment vector caused by the weight, the propulsion mechanism and the spherical shell respectively, and I_j^c, I^s are both constant. In Eq. (8), I_{jxx}^c and I_{jyy}^c are as follows:

$$\begin{aligned} I_{jxx}^w &= \bar{I}_{jxx}^w + mdx^2, \\ I_{jyy}^w &= \bar{I}_{jyy}^w + mdy^2, \\ I_{jzz}^w &= \bar{I}_{jzz}^w + mdz^2 \end{aligned} \quad (9)$$

Equation (9) is transfer-of-axis relations for transferring \bar{I} from the center of mass to Os. Also, in Eq. (9), due to symmetrical design of the robot $I_{xy}^c = I_{yx}^c$, and the other parameters are calculated likewise.

Let X, Y and Z be the projection of the weight on the body coordinate axes be the angular velocity of the sphere, the relational equation can be expressed by

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \left(\prod_{k=n-1}^0 A_k \right) \begin{bmatrix} 0 \\ -mg \\ 0 \end{bmatrix} \quad (10)$$

In the above equation, the elements of the ration matrix A_k is given by

$$A_k = \begin{bmatrix} c\beta_k c\gamma_k & c\alpha_k s\gamma_k + s\alpha_k s\beta_k c\gamma_k & s\alpha_k s\gamma_k - c\alpha_k s\beta_k c\gamma_k \\ -c\beta_k s\gamma_k & c\alpha_k c\gamma_k - s\alpha_k s\beta_k s\gamma_k & s\alpha_k c\gamma_k + c\alpha_k s\beta_k s\gamma_k \\ s\beta_k & -s\alpha_k c\beta_k & c\alpha_k c\beta_k \end{bmatrix} \quad (11)$$

We assume that $\varepsilon_x^n, \varepsilon_y^n$ and ε_z^n are the projection of the angular velocity of the sphere in the $(n-1)^{th}$ step, given by

$$\begin{bmatrix} \varepsilon_x^n \\ \varepsilon_y^n \\ \varepsilon_z^n \end{bmatrix} = \begin{bmatrix} \dot{\omega}_\xi^n \\ \dot{\omega}_\eta^n \\ \dot{\omega}_\zeta^n \end{bmatrix} = \begin{bmatrix} 1 & 0 & s\beta_n \\ 0 & c\alpha_n & -s\alpha_n c\beta_n \\ 0 & s\alpha_n & c\alpha_n c\beta_n \end{bmatrix} \begin{bmatrix} \ddot{\alpha}_n \\ \ddot{\beta}_n \\ \ddot{\gamma}_n \end{bmatrix} \quad (12)$$

$$+ \begin{bmatrix} 0 & 0 & \dot{\beta}_n c\beta_n \\ 0 & -\dot{\alpha}_n s\alpha_n & -\dot{\alpha}_n c\alpha_n c\beta_n + \dot{\beta}_n s\alpha_n s\beta_n \\ 0 & \dot{\alpha}_n c\alpha_n & -\dot{\alpha}_n s\alpha_n c\beta_n - \dot{\beta}_n c\alpha_n s\beta_n \end{bmatrix} \begin{bmatrix} \dot{\alpha}_n \\ \dot{\beta}_n \\ \dot{\gamma}_n \end{bmatrix}$$

From Eq. (8), Eq. (10) and Eq. (12) the resultant moment Equation in the n^{th} step is as following

$$\begin{bmatrix} Z \sum_{K=1}^4 y_{k-n} - Y \sum_{K=1}^4 z_{k-n} \\ X \sum_{K=1}^4 z_{k-n} - Z \sum_{K=1}^4 x_{k-n} \\ Y \sum_{K=1}^4 x_{k-n} - X \sum_{K=1}^4 y_{k-n} \end{bmatrix} = \left[\sum_{K=1}^4 (I_k^w + I_k^c) + I^s \right] \begin{bmatrix} \varepsilon_x^n \\ \varepsilon_y^n \\ \varepsilon_z^n \end{bmatrix} \quad (13)$$

Therefore, by Eq. (5), Eq. (6) and Eq. (13), the next position of weights can be calculated as below:

$$\begin{bmatrix} Z \sum_{K=1}^4 y_{k-n} - Y \sum_{K=1}^4 z_{k-n} \\ X \sum_{K=1}^4 z_{k-n} - Z \sum_{K=1}^4 x_{k-n} \\ Y \sum_{K=1}^4 x_{k-n} - X \sum_{K=1}^4 y_{k-n} \end{bmatrix} = \left[\sum_{K=1}^4 (I_k^w + I_k^c) + I^s \right] \begin{bmatrix} \varepsilon_x^n \\ \varepsilon_y^n \\ \varepsilon_z^n \end{bmatrix} \quad (14)$$

$$\dot{x}_n + (\dot{\beta}_n s\alpha_n + \dot{\gamma}_n c\alpha_n c\beta_n) r = 0$$

$$\dot{z}_n - (\dot{\alpha}_n + s\alpha_n + \dot{\gamma}_n s\beta_n) r = 0$$

With the Eq. (14), we can calculate all possible positions of weights.

6. Simulation and Experiments

ADAMS is the most widely used multi-body dynamics and motion analysis software in the world. Traditional “build and test” design methods are now too expensive, too time consuming, and sometimes even impossible to do. ADAMS multi-body dynamics software enables engineers to easily create and test virtual prototypes of mechanical systems in a fraction of the time and cost required for physical build and test. Unlike most CAD embedded tools, ADAMS incorporates real physics by simultaneously solving equations for kinematics, static, quasi-static, and dynamics.

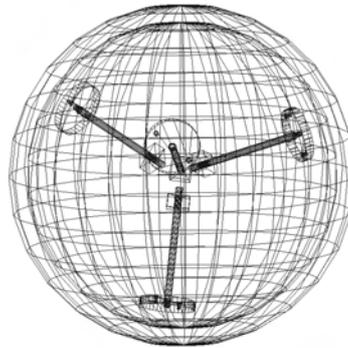


Fig. 4. 3D model of SMR in ADAMS

Due of 3D solid modeling ability of ADAMS is not very strong, we employed the SOLIDWORKS2007 software to set up the 3D model, and then export to ADAMS2007 for simulation, the model is shown in Figure 4. Also, because the ADAMS' controllers are not ones for professional modeling tool, so its control capabilities are limited, however, interfaces between ADAMS and other soft wares such as MATLAB, EASY and so on allow ADAMS complete fine system simulation.

In this section, simulation results on the spherical robot by ADAMS using a time step 0.01s are presented to demonstrate the effectiveness of the design and verify the path following performance. The system parameters are given in Table 1, and the experiment results shown in Fig. 5 and Fig. 6.

The evolutions of the variables for tracking the desired curve are shown in Fig.5 (a, b). The reference linear velocity of the weight v_r is 100mm/s. Fig. 5 (a, b) indicate that the robot starts roll in the x-z plane at $t=0.75s$, in contrast with Fig.5 (c, d) where the robot starts at $t=0.5s$. This corresponds to a pure rolling motion. Simulation indicates that the angular velocities generated oscillatory and chattering behavior with these assumptions.

Modeling and Simulation of a Spherical Mobile Robot

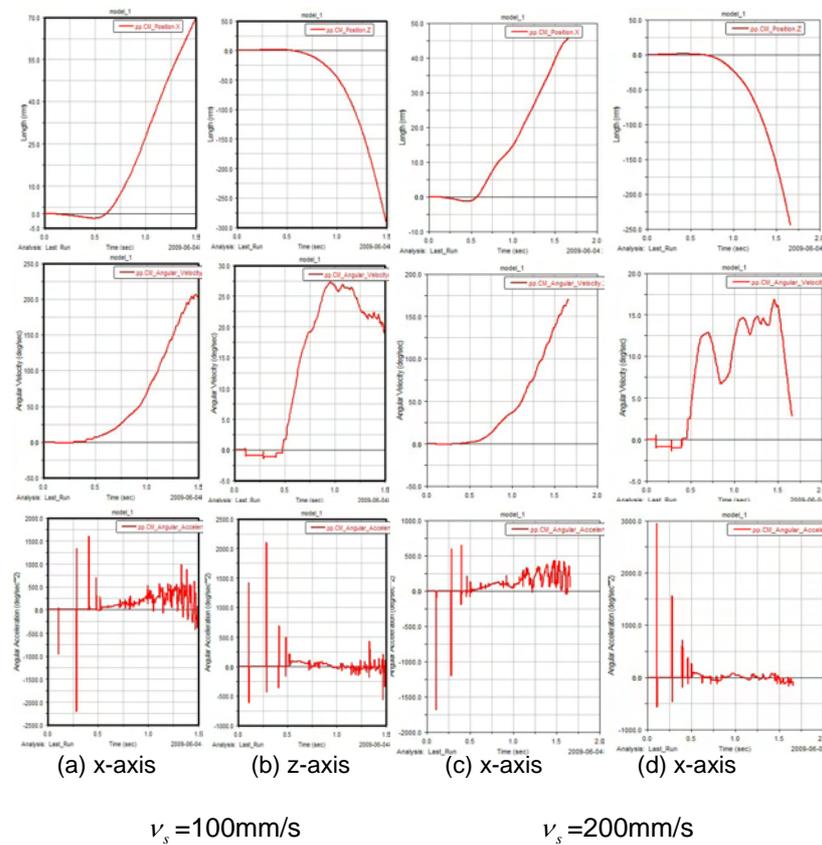


Fig. 5. The performance of SMR at $v_s = 100\text{mm/s}$

Table 1. The experimental setup configuration of SMR

Variable	Quantity	Unit
Holding torque of motors	3.250	kg/m
Weight of motor (each)	0.550	m
Radius of spokes	0.040	m
Pitch of power screw	0.060	m
Weight of power screw (each)	0.206	kg
Weight of unbalanced masses (each)	1.125	kg
Radius of spherical shell	0.300	m
Weight of spherical shell	3.779	kg

Fig.6 shows the results of the robot in tracking a straight line in x-z plane, in contrast with reference line (red). As expected, simulations reveal that the spherical robot converges globally uniformly to the desired point with acceptable dynamic performances.

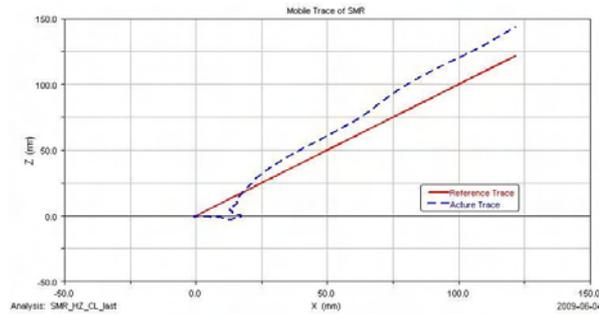


Fig. 6. The trajectories of SMR on the x-z plane

For several missions, the experimental results agree well with those of the simulations. In each case, the experimental trajectory follows the predicted one with a reasonable accuracy. Some factors contribute to these inaccuracies are:

- 1) the center of mass of the robot is not exactly at the geometric center of the robot;
- 2) imperfections on the surface of the sphere;
- 3) open-loop nature of the robot control.

The following are several issues of experimental results. The figures are in three parts; figure in the top-left shows the distance between the robot's center and origin, where the robot has began its travel, figure in bottom-left shows the absolute velocity of robot and figure on the right shows the movement of robot in x-z plane, predicted by dynamic model.

7. Conclusions

A mathematical model of omnidirectional spherical mobile robot motion was established using the no-slip rolling constraint and conservation of angular momentum, and an algorithmic motion planning was developed. The model was validated through a set of simulations. Results of simulations and experimental trajectories of the robot on the plane were found consistent with a reasonable accuracy and the methods are effective. Trajectories are quite accurate despite lack of on-board feedback control. Comparing with existing motion plans, most of which require intensive numerical calculations, strategies in this paper involve simple algorithmic iteration and provide the scope for easy implementation. However, experiments also show that the

spherical robot has a strong tendency to oscillate and the uneven ground could make the robot oscillate for a long time. So the control problem of spherical robot couldn't be solved by open-loop control strategy and a robust closed-loop controller and suitable stabilization method are necessary. Study in this paper demonstrates the feasibility of the approaches and a better controlled robot is expected to be improved in the future.

8. Acknowledgments

The work was partially supported by the Natural Science Foundation of Shandong, China (2009ZRA01105), the science and technology development fund of Tai'an (20083002) and the Foundation provided by Taishan College (Y06-2-15).

9. References

1. A. Halme, T. Schonberg, Y. Wang, Motion control of a spherical mobile robot, in 4th IEEE International Workshop on Advanced Motion Control AMC'96, Mie University, Japan, 1996, pp. 100-106.
2. F. Michaud, S. Caron, Roball: the rolling robot, *Autonomous Robots*, vol. 12, no. 2, 2002, pp. 211-222.
3. A. Bicchi, A. Balluchi, Introducing the 'SPHERICLE': an experimental testbed for research and teaching in nonholonomy, in *Proceedings of the 1997 IEEE Int. Conf. Robotics and Automation*, Albuquerque, New Mexico, April 1995, pp. 2620-2625.
4. A. Javadi, P. Mojabi., Introducing august: a novel strategy for an omnidirectional spherical Rolling Robot, in *Proc. IEEE Int. Conf. Robotics and Automation*, 2002.
5. Qiang Zhan, Tingzhi Zhou, Ming Chen, Sanlong Cai. Dynamic Trajectory Planning of a Spherical Mobile Robot. in *Proc. IEEE Int. Conference. Robotics Robotics, Automation and Mechatronics*, 2006, pp. 1-6
6. Masayoshi, W., and Shunji, M., "Modeling and Control of a New Type of Omnidirectional Holonomic Vehicle," in *AMC896*, 1996, pp. 265-270.
7. "Wheels," in *Proceedings of IEEE International Conference on Robotics and Automation*, 1996, pp. 3664-3670.
8. Tang, J., Watanabe, K., and Shiraishi, Y., "Design and Traveling Experiment of an Omnidirectional Holonomic Mobile Robot," in *Proc. IEEE Int. Conf. Intelligent Robots and Sys.*, 1996, pp. 66-73.
9. Agrachev, A. A., "An Intrinsic Approach to the Control of Rolling Bodies," in *Proceedings of the Conference on Decision and Control*, 1999, pp. 431-435.
10. Bhattacharya, S., and Agrawal, S. K., "Spherical Rolling Robot: A Design and Motion Planning Studies," in *Proceedings of IEEE International Conference on Robotics and Automation*. 2000,
11. Mukherjee, R., Minor, M. A., and Pukrushpan, J. T., "Simple Motion Planning Strategies for Spherobot: A Spherical Mobile Robot," in *Proceedings IEEE International Conference on Decision and Control*, 1999, pp. 2132-2137.
12. T. Ylikorpi, J. Suomela, Ball shaped robots: An historical overview and recent development at TKK, *Field and Service Robotics*, vol. 25, no. 6, 2006, pp. 343-354.

SANG Shengju, ZHAO Jichao, WU Hao, CHEN Shoujun, and AN Qi

13. R.H. Armour, J.F.V. Vincent, Rolling in nature and robotics: A review, *Journal of Bionic Engineering* 3, 2006, 195–208.
14. T. Das, R. Mukherjee, Exponential stabilization of the rolling sphere, *Automatica*, 2004, 1877–1889.
15. T. Das, R. Mukherjee, Reconfiguration of a rolling sphere: A problem in involute–involute geometry, *ASME Journal of Applied Mechanics*, 2006, 590–597.
16. Fredrik C. Bruhna, Henrik Kratza, JohanWarellb, etc. Apreliminary design for a spherical inflatable microrover for planetary exploration. *Acta Astronautica*, vol. 63, 2008, pp.618 – 631.
17. A. Gilyén and P. Szvoboda. Development of the Hungaroshere: the Husar-11 Rover within a Transparent Sperial Space Probe Model with Special Planetary Surface Activities. 40th Lunar and Planetary Science Conference, 2009
18. Stefan Staicu. Recursive Modelling in Dynamics of Agile Wrist Spherical Parallel Robot. *Robotics and Computer-Integrated Manufacturing*, vol. 25, 2009, pp. 409-416.
19. Zhan Qiang, Liu Zengbo and Cai Yao. A Back-stepping Based Trajectory Tracking Controller for a Non-chained Nonholonomic Spherical Robot. *Chinese Journal of Aeronautics*, vol. 21, 2008, pp. 472-480.
20. V. Joshi, R. Banavar, Motion analysis of a spherical mobile robot, *Robotica*, 2009, 343-353, doi: 10.1017/ S0263574708004748.
21. V.A. Joshi et al., Design and analysis of a spherical mobile robot, *Mech. Mach. Theory*, 2009, doi: 10.1016/ j.mechmachtheory.2009.04.003

Sang Shengju was born in Taian, P.R.China in 1965. He received his M.Sc degree in Measurement-Control Tech from Harbin University of Science and Technology, Harbin,P.R.China in 1991. He is currently a Ph.D Candidate in the East China University of Science and Technology, Shanghai, P.R.China. and working as a senior Engineer of computer science at theTaishan College, Taian, P.R.China. His research interests include image processing, embedded system and spherical robot.

Zhao Jichao is a professor of Computer Science at the Taishan College, Taian, P.R.China. His current research focuses on image processing, embedded system and spherical robot.

Chen Shoujun and **Wu Hao** are Ph.D students in the School of Mechanical and Power Engineering, East China University of Science and Technology. The areas of interest and research are galling mechanism and preventive methods with anti-sulfur oil tubing in high steel grade.

An Qi received his PhD degrees from the Department of Mechanical Engineering, Zhejiang University, Hangzhou, P.R.China in 2001.He is currently a full professor, Ph.D tutor in the School of Mechanical and Power Engineering at the East China University of Science and Technology, Shanghai, P.R.China. His research interests include Image Processing and CAD/CAPP/CAM.

Received: May 06, 2009; Accepted: October 15, 2009.

3D Mesh Skeleton Extraction Using Prominent Segmentation

Xiaopeng SUN¹, J. PAN², and Xiaopeng WEI³

¹ School of Mechanical & Engineering, Dalian University of Technology,
Dalian 116024, China
xpzhSun@gmail.com

² Department of Computer & Information Technology,
Liaoning Normal University, Dalian 116029, China
cadcg2008@gmail.com

³ Key Laboratory of Advanced Design and Intelligent Computing (Dalian University),
Ministry of Education, Dalian 116622, China
xiaopengwei@gmail.com

Abstract. Skeleton of 3D mesh is a fundamental shape feature, and is useful for shape description and other many applications in 3D Digital Geometry Processing. This paper presents a novel skeleton extraction algorithm based on feature point and core extraction by the Multi-dimensional scaling (MDS) transformation. The algorithm first straightens the folded prominent branch up, as well as the prominent shape feature points of mesh are computed, a meaningful segmentation is applied under the direction of feature points. The Node-ring of all segmented components is defined by discrete geodesic path on mesh surface, and then the skeleton of every segmented component is defined as the link of the Node-ring's center. As to the core component without prominent feature points, principal curve is used to fit its skeleton. Our algorithm is simple, and invariant both to the pose of the mesh and to the different proportions of model's components.

Keywords: Skeleton, MDS, Discrete geodesic path, Node-ring.

1. Introduction

In recent years, how to design a simple and robust algorithm to extract the skeleton of object with less memory cost and shape information loss has become a fundamental problem of information visualization and pattern recognition [1] et al.

Most existing algorithms were developed for 3D medical image analysis with volumetric data as input, and little work on the 3D mesh models, which are represented as a polygonal list. But because of the property of reducing the dimension of the problem, skeleton has shown its promising advantage in the application of geometry and topology description with its simplicity in region of 3D Digital Geometry Processing (DGP), such as deformation[2][3],

shape recognition and retrieval[4][5], mesh editing[6], simplification[7], motion control and collision detection[8], segmentation[9] et al.

Meaningful mesh segmentation is another fundamental problem in DGP, which decompose 3D surface meshes into functional shape components, not only provides semantic information about the underlying mesh, but also could be used to guide several mesh processing algorithms, including skeleton extraction, modeling, morphing, shape-based 3D shape retrieval, and texture mapping[9] et al. All of these applications benefit from mesh meaningful segmentations that obey human intuition. Good 3D mesh segmentation will result in high quality skeleton.

In this paper, we introduce a novel skeleton extraction algorithm based on meaningful shape decomposition, which segment the 3D mesh under the direction of the prominent feature point and core. The challenge of this algorithm is that, the triangle patch size should be uniform in general level to keep a high precision, because the Node-ring is defined by the length of discrete geodesic path, or subdivision must be applied firstly. And to those meshes with 5,000 vertices or more, efficient memory management should be considered to improve computing speed of shortest path between all pair vertices of mesh, or simplification must be applied as a preprocessing. Due to the transformation of MDS and the direction of prominent feature points, we decompose the mesh into several prominent meaningful branch components and a core component, and then compute the skeleton of meaningful components respectively, so our algorithm is robust, simple, and pose-invariance.

2. Related works

Cornea et al. presented a comprehensive overview on curve-skeleton properties, applications and algorithms[1]. In recent decades, hundreds research works have been published in very extensive application regions.

The skeleton model is initially defined as a collection of the centre of the largest inscribed sphere[10], and is used extensively in virtual navigation, traditional computer graphics, medical image segmentation and quantification, registration, matching, 3D mesh morphing, segmentation, and analysis of scientific data et al.

And the various applications required their own desired properties of the skeleton; for example, the skeleton should be topologically equivalent to the original object in shape recognition, and must be invariant under isometric transformations[4][11]. In the application of shape compression and volume animation, the skeleton should provide enough information to reconstruct a 3D object completely from its medial skeleton representation by computing the union of maximal inscribed balls[12]. And the skeleton should be its centeredness within the 3D object[13]; the logical components of the object should have a one-to-one correspondence with the logical components of the skeleton[14][15]; be not very sensitive to little noise in the boundary;

approximate to the complex components of an object, and reflect the natural hierarchy of these complexities, et al.

Commonly, the curve-skeleton algorithms can be divided into three classes: topological thinning (grassfire propagation), distance transform based (ridge detection) and Voronoi diagram based[16][17]. More recently in 2007, Cornea et al. initially categorize the algorithms into four new classes: thinning and boundary propagation, using a distance field, geometric methods, general field functions[1]. Our algorithm is a geometric method.

The basic idea of the thinning algorithm to extract the skeleton of voxel model is: from outside to inside, layer upon layer strips[18]. To judge one voxel whether needed to be stripped off is comparatively time-consuming work, therefore, G. Bertrand proposes a parallel thinning algorithm[19]. However, many problems are still not solved well at present, such as some skeletons are not continuous and affected by noise, often contain some invalid branches. And skeleton with a higher topology and centrality is very hard to extract et al.

Skeleton extraction based on the distance transformation algorithm can ensure the centrality of skeleton points, but is weak in keeping original topology. The algorithm based on the Level Set[20] has a better stability and higher topology independency, can effectively overcome cusp and skeleton fracture, but even Fast Marching Method is used, its computational complexity is still very high. A snake model could adjust skeleton position, to improve skeleton's centrality, but increased complexity as well.

Compared with the distance transform and thinning algorithms, the algorithm based on shape classification or segmentation has a more remarkable superiority [21][22]. Based on the topological connection information, segmentation is applied firstly; then the skeleton is extracted from the segmentation results, to reduce the complexity and enhance skeleton's precision.

On the other hand, the 3D skeleton can be used to guide segmentation in turn, to get a more meaningful segmentation and obey the Minimal Rule better[9][23]. Most skeleton extraction works involved with segmentation are sensitive to the pose of the model. For instance, very different segmentations will be produced when the models of human have their arms folded or not, due to the vital difference of curvature (or dihedral angles). By decomposing a 3D model into approximate convex components, Lien etc calculated the vertices convexity-concavity firstly, and then extracted the skeleton by iteration, but the obtained skeleton had a worse centrality, and hard to establish the level relations[24].

Segmentation is a classical problem in processing of 3D mesh surfaces, and other types of multimedia data, and most of these methods have been evaluated only by visual inspection of results, reference [23] provide four quantitative metrics for evaluation of mesh segmentation algorithms, that are cut discrepancy, Hamming distance, rand index, and consistency error. The focus of the segmentation in this paper is to find the prominent branch and improve the speed, so we neglect the above four quantitative metrics, and leave the jaggy boundaries to the feature.

3. Overview

Inspired by the work of Sagi Katz etc [25], we proposed a novel skeleton extract algorithm, which based on segmentation of prominent branches and geodesic path.

Given an 3D mesh S in Euclid space, its vertices set is $\{v_i | 1 \leq i \leq n\}$, and $\bar{\delta}_{ij} = \text{GeoDist}(v_i, v_j)$ is defined as the weighted discrete geodesic distance between every two adjacent vertex (v_i, v_j) on S [14]. Based on Fast Marching method in [26], to every pair of vertices (v_i, v_j) on S , the discrete geodesic distance $\bar{\delta}_{ij} = \text{GeoDist}(v_i, v_j)$ and the shortest path $\text{ShortPath}(v_i, v_j)$ can be obtained.

Our algorithm has the following steps:

Based on the theory of multi-dimensional scaling (MDS), the mesh S is transformed from Euclid space into its pose-invariant representation S_{MDS} in MDS space, and the folded organs branches are prominently straightened.

Robust prominent feature points $PF_i, i=1,2,\dots,k$ can be located on the convex-hull in MDS space, and then S is segmented into several prominent meaningful branch components S_1, S_2, \dots, S_k and a core component S_0 .

Define the points located on the boundary between the components as **connecting ring** $LOOP_i$.

Compute the discrete geodesic path from PF_i to every point on $LOOP_i$, the space structure of there shortest paths can be represented by a tree with its root on the feature point.

From the root PF_i to the $LOOP_i$, we define 20-30 level sets of mesh points, called **node ring**.

Fitting the connect line of the center of node ring with KS principal curve, skeletons of branches are computed. And the skeleton of core is computed specially.

Finally, joint the branch skeleton of every component together in a simple way, the single connected skeleton of S obtained.

Every sub-mesh's node ring is defined by the discrete geodesic distance on mesh surface, and the skeleton of each sub-mesh is extracted, our algorithm can deal with bending posture and overcome of noise, and is simple, robust, and pose-invariant.

4. Segmentation of prominent brances

Due to noise, it is difficult to extract feature points on mesh model S directly. The MDS (Multi-dimensional scaling) is used to transform S into the mesh S_{MDS} in MDS space, and let d_{ij} be the Euclidean distance between the corresponding points of (v_i, v_j) on S_{MDS} , and then the distances matrix is obtained.

After 30-50 iteration optimizations by the stress function in [25], the topology information of model S is kept, but the position of every vertex is

changed, every folded prominent components of a model are straightened, we denoted the optimization process by $S \rightarrow S_{MDS}$.

Based on MDS transformation, $S \rightarrow S_{MDS}$ can filter the noise on model S globally, and obtained a new mesh S_{MDS} with every folded prominent branch straightened. We extract the prominent feature points on S_{MDS} to overcome the interference of noise.

In other words, if a vertex be a prominent feature point on S , it should reside on a tip of S_{MDS} , (i.e. on the convex-hull of S_{MDS}), and it is a local maximum of the sum of the geodesic distance to any other vertex[25].

We denote the k prominent feature points on S_{MDS} by $PF_i, i=1,2,\dots,k$ (e.g. the red points in Fig. 1)



Fig. 1. Feature points (in red).

Then we define a mirror sphere on S_{MDS} , so every vertex v of S_{MDS} has a image v_{mirror} outside the mirror sphere[25]. Define the core of mesh S_{MDS} as a set of the vertices reside on the convex-hull, and denoted it by SC_{MDS} . Obviously, the prominent branches and feature points of S_{MDS} become the internal of the convex-hull.

Let the number of feature points PF_i on S_{MDS} be k , so its corresponding prominent branches number is k too. After the mirror transformation, the prominent branches will be cut by the core component extraction, then we obtain $k+1$ meaningful segmentation components $S_0, S_1, S_2, \dots, S_k$ of mesh S , where S_0 is the core component corresponding to SC_{MDS} (e.g. the green component in Fig. 2-(e)), and $S_i, i=1,2,\dots,k$ is sub-mesh composed of k prominent branches (e.g. the white, pink, red and yellow components in Fig. 2-(e)).

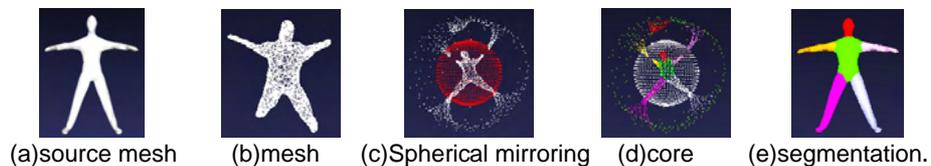


Fig. 2. Spherical mirroring segmentation

5. Construction of shortest path and skeleton

5.1. Connecting ring

Let $S_0, S_1, S_2, \dots, S_k$ be the $k+1$ segmentation of model S , and denoted the vertex set on the common boundary between sub-mesh $S_i, i=1, 2, \dots, k$ and S_0 by $LOOP_i = \delta S_0 \cap \delta S_i$ (e.g. the red points in Fig. 3-(a)), then named this vertex set as a **connecting ring** between every sub-mesh S_i and the core component S_0 .

For every shortest path between prominent feature point PF_i and the vertex in the corresponding connecting link $LOOP_i$, obviously the number of such shortest path is more than one, and the space structure of these shortest paths can be represented by a tree whose root is the feature point PF_i , we denote this tree as $Tree_i$.

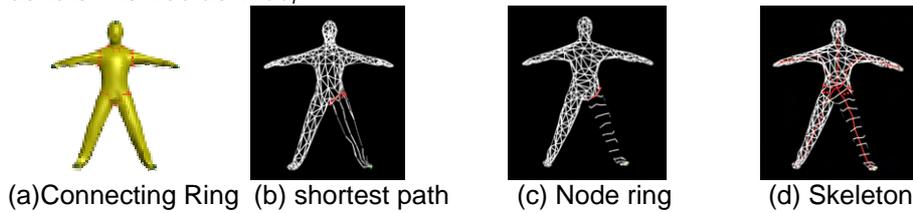


Fig. 3. Connecting Ring and skeleton of branch

5.2. Node ring of S_i and skeleton computing

For the shortest path tree $Tree_i$ with the feature point PF_i as root in the sub-mesh S_i , the depth from the root node PF_i to each leaf node on corresponding connecting ring $LOOP_i$ is different, and corresponding discrete geodesic distances is different too.

The shortest path set of feature point PF_i on sub-mesh S_i to any v_j on $LOOP_i$ is denoted by $ShortPath(PF_i, v_j)$ (e.g. Fig. 3-(b)), the geodesic length of corresponding shortest path is denoted by $sp(v_j) = || ShortPath(PF_i, v_j) ||$, and the maximum and the minimum length of all the discrete geodesic distance on this set are denoted by $sp_{max} = \max\{|| ShortPath(PF_i, v_j) ||\}$ and $sp_{min} = \min\{|| ShortPath(PF_i, v_j) ||\}$ respectively.

Let $\angle sp = sp_{max} / m$, $\delta sp = sp_{min} / m$, then define the m **node rings** (e.g. the level ring within the right leg in Fig. 3-(c)) of sub-mesh S_i as its m subsets of vertex set $H_t(v) = \cup \{v \in S_i | \delta sp \cdot t \leq sp(v_j) \leq \angle sp \cdot t\}$, where m is a nonnegative constant integer, let $t=1, 2, \dots, m$, and $C = || H_t(v) ||$, then the center coordinates of node ring $H_t(v)$ be defined as:

$$Center(H_t(v)) = \sum_{c=1}^c \frac{Coordinate(H_t(v))}{C}$$

Connect the centroid $Center(H_t(v))$ of successive level sets on sub-mesh S_i , and the center of $LOOP_i$, then the skeleton curve of sub-mesh S_i be constructed by the connected $m+1$ points (e.g. the red line within the right leg in Fig. 3-(d)).

Do the same processing to all sub-mesh S_1, S_2, \dots, S_k , we will get the skeleton of each prominent branch except S_0 , which corresponds with the core component.

5.3. Skeleton of core components S_0

Due to the core component S_0 has no feature points, so a special consideration is needed to extract its skeleton.

For the core component S_0 with its gravity center located in internal, the KS principal curve algorithm in [27] is used to fit k lines which connected the central point of $LOOP_i$ and the center of core component S_0 , the new skeleton S_0^{ks} is represented by principal curve. Connect the central point of $LOOP_i$ and relative nearer endpoint of S_0^{ks} by turn, the complete skeleton of model S is obtained.

For the core component S_0 with gravity center located in exterior, its skeleton could be obtained by following method:

For given S and its vertex set $\{v_i | 1 \leq i \leq n\}$ with coordinate set $\{x_i, y_i, z_i\}_{i=1}^n$, the moment (p, q, r) -th of discrete definition can be approximated by

$$m_{pqr} = \frac{1}{n} \sum_{i=1}^n x_i^p y_i^q z_i^r$$

According to its definition, the set of moments $\{m_{pqr}\}$ are uniquely, and are uniquely determined by the object, the first moments and second moments respectively represent the mesh's center of the mass and the three PCA axes [28]. Then, based on the first moment and second moment of the core component S_0 , the gravity center, the PCA axes and three axis planes of core component S_0 can be obtained.

Let O be the gravity center of S_0 , and sort the three main axes according to the axial length in descending order as A, B, C , let the base plane be OBC where the two shorter axis lie in, a constant h_0 is suitably chosen along OA , i.e. the longest axis direction. Horizontally cut the vertex set S_0 into m subsets with thickness h_0 , denote the vertex subsets of S_0 by $H_t^0(v)$, $t=1, 2, \dots, m$, named $H_t^0(v)$ by the **central ring** of the core component S_0 , with the central coordinate as

$$Center(H_i^0(v)) = \sum_{c=1}^{c^0} \frac{Coordinate(H_i^0(v))}{C^0}$$

Where $C^0 = ||H_i^0(v)||$. Connect $Center(H_i^0(v))$ sequentially; the skeletal curve of core component S_o will be a ks principal curve fitting of the lines by the m connecting points.

Our algorithm requests core component has the better convexity, then $Center(H_i^0(v))$ should locate within the projecting polygon of the vertex subset $H_i^0(v)$ on base plane OBC , otherwise, the skeleton will go out the core component S_o , and be not located in the internal of model S completely, this algorithm expiration.

6. Results and discussion

We use the Shape Benchmark database at Princeton university (Fig. 4-(h) exceptions) to test our algorithm. All data are recorded on an 2.20GHz Intel Core(TM) 2 Duo CPU E4500 machine with 2GB RAM and 128M GM, using a single thread implementation.

Table 1 shows the running time of computing the prominent feature point and skeleton for various 3D models.

Table 1. Computing time

Model	Vertices	Faces	Feature Points	Time(s)
M0235	408	812	5	0.311
M0336	1,557	3,110	6	3.813
M0324	1,564	3,124	6	4.001
M0031	1,802	3,548	5	4.876
M0178	914	1,824	5	1.871
M0071	420	836	6	0.345
M0095	956	1,908	6	2.062
M0000	3,117	6,250	9	14.794
M0050	1,455	2,906	3	3.304
M0058	477	950	5	0.502
M0104	906	1,808	7	1.831

In Fig. 4, the green dots are the prominent feature points, yellow points are two endpoints of the core component skeleton, and red lines are the skeletons of its core and branches.

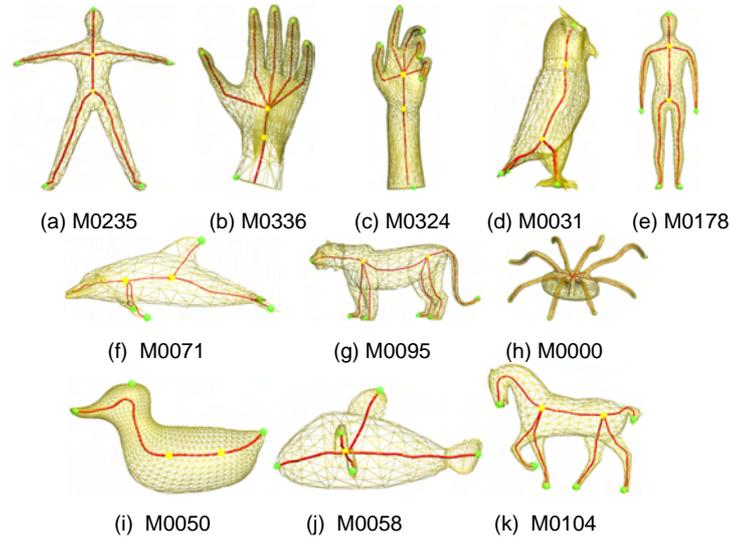


Fig. 4. Skeletons of some models

Although the optimization process $S \rightarrow S_{MDS}$ stretching the mesh S , to help the extraction of prominent feature points on the pose-invariant model S_{MDS} , can overcome the interference of noise, but on the other hand, some slight partial characteristic be shielded, as shown by owl's beak in Fig. 4-(d), and ears in Fig. 4-(g) and Fig. 4-(k) etc, where the smaller feature points are shielded by the more prominent neighboring extreme points. In addition, to those models without prominent branches, the obtained skeleton does not have clear meaning, just as Fig. 4-(h) ellipsoid part shows.

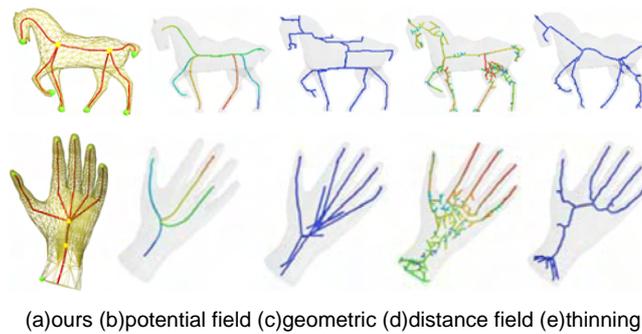


Fig. 5. Skeleton obtained using different algorithms

The skeleton of the prominent branch start not from the feature point PF_i , but from the centroids $Center(H_i(v))$ of successive level sets of S_i , until arrive the center of $LOOP_i$, as Fig. 4-(b) and Fig. 4-(c) shows, skeleton starts away from the green feature points at the bottom of the wrist; and as a fact, the

green feature point and the prominent branch skeleton's red endpoint are also separate in other results.

Fig. 5 shows the results of our algorithm and the algorithm of thinning distance field, geometric, potential field method. The left column is produced by our algorithm, and other four columns are provided by [1]. Obviously, the comparison shows that our algorithm is more effective than others.

Let N be the number of faces in the original model S , then the computational complexity of MDS is bounded by $O(N^2)$; the prominent feature point extraction takes $O(\text{Mog}N)$; the $k+1$ segmentation costs $O(\text{Mog}N)$; connecting ring, node ring and center ring costs $O(\log N+N)$; skeleton fitting costs $O(N)$; thus the overall time complexity is $O(N^2+\text{Mog}N+N+\log N)=O(N^2)$.

In addition, our algorithm can extract the skeleton of simply connected and complex topology 3D model with its genus <1 . For constructing by connecting prominent feature points PF_i and the points on connecting ring $LOOP_i$, the pyramidal shortest path set guarantees the skeleton to locate in the internal of sub-mesh S_i , and enhances the centrality of skeleton.

7. Conclusion and feature work

We proposed a novel algorithm for 3D mesh skeleton extraction. The advantage of this approach is that: Our algorithm does not involve complicated mathematical tools. Compared with snake and level set[20], our algorithm has a lower complexity. It obtains the prominent feature points by the preprocessing of MDS transformation, and then extracts the local skeleton under the guidance of the prominent feature points and the strategy of segmentation, reduces the influence of noise, and improves the efficiency. The skeleton obtained is topology-invariant and pose-invariant. The shortest path's search is carried on the sub-mesh, overcomes the difficulty to judge branches and select feature points; our algorithm requires neither prior-knowledge of the number of feature points nor any interactive user parameters; and our skeleton is invariant to the pose of the mesh model.

The limitations of our algorithm are: When the core component has prominent concavity, its gravity center located outside the projection of core component on base plane OBC, skeleton will go out the surface. In addition, we connected the skeleton of core component and that of other prominent branches without trimming strategy, then the skeleton was not guaranteed to be smoothing in any case. And obviously, the prominent feature points of sub-mesh S_i used as the root of shortest path tree can be used to roughly determine skeleton line shape and trends on the consideration of its relative position to connecting ring, this potential information was not fully used for now.

8. Acknowledgment

This work is supported in part by National Natural Science Foundation of China with projects No. 60873110, No. 60875046, and No. 60533090.

9. References

1. Cornea, N. D., Min, P.: Curve-skeleton properties, applications, and algorithms. *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, No. 3, 530–548. (2007)
2. Oscar Kin-Chung Au, Chiew-Lan Tai, Hung-Kuo Chu, Daniel Cohen-Or, Tong-Yee Lee: Skeleton Extraction by Mesh Contraction. *ACM Transactions on Graphics (SIGGRAPH 2008 issue)*, Vol. 27, No. 3, 44:1-44:10. (2008)
3. Han-Bing Yan, Shi-Min Hu, Ralph R Martin, and Yong-Liang Yang: Shape Deformation using a Skeleton to Drive Simplex Transformations. *IEEE Transaction on Visualization and Computer Graphics*, vol. 14, No. 3, 693-706. (2008)
4. Masaki Hilaga, Yoshihisa Shinagawa, Taku Kohmura, Tosiyasu L. Kunii: Topology Matching for Fully Automatic Similarity Estimation of 3D Shapes. In *Proceedings of ACM SIGGRAPH*. Los Angeles, USA, 203-212. (2001)
5. Sundar, H., Silver, D., Gagvani, N., Dickinson S.: Skeleton based shape matching and retrieval. In *Proceedings of International Conference on Shape Modeling and Applications*, Seoul, Korea, 130-139. (2003)
6. Tao Ju, Qian-Yi Zhou, Shi-Min Hu: Editing The Topology of 3D Models by Sketching. *ACM SIGGRAPH 2007, ACM Transactions on Graphics*, vol. 26, No. 3, 42:1-42:10. (2007)
7. Tam, R., Heidrich W.: Shape simplifications based on the medial axis transform. In *Proceedings of the 14th IEEE Visualization*, Seattle, WA, USA, 63-70. (2003)
8. Jianfei Liu, Xiaopeng Zhang, Qingqiong Deng: Structural Volume Skeletonization and Visibility Computation for Virtual Endoscope. *Journal of Computer-Aided Design & Computer Graphics*, vol. 19, 1352-1358. (2007)
9. Xiaopeng Sun, Hua Li: A Survey of 3D Mesh Model Segmentation and Application. *Journal of Computer-Aided Design & Computer Graphics*, vol. 17, No. 8, 1647-1655. (2005)
10. Blum H.: A Transformation for Extracting New Descriptors of Models for the Perception of Speech and Visual Form. W. Walthen-Dunn, ed., US: MIT Press, Combridge. (1967)
11. Cornea N.D., Demirci M.F., Silver D., Shokoufandeh A., Dickinson S. J., Kantor P. B.: 3D Object Retrieval using Many-to-many Matching of Curve Skeletons, In *Proceedings of Shape Modeling and Applications*, Cambridge, MA, USA. (2005)
12. Gagvani, N., Silver D.: Animating volumetric models. *Academic Press Professional*, vol. 63, No. 6, 443-458. (2001)
13. Dey, T.K., Sun J.: Defining and Computing Curve-Skeletons with Medial Geodesic Function, In *Proceedings of Eurographics Symp. Geometry Processing*, Cagliari, Sardinia, Italy, 143-152. (2006)
14. Sagi Katz, Ayellet Tal: Hierarchical Mesh Decomposition using Fuzzy Clustering and Cuts. *ACM Transactions on Graphics*, vol. 22, No. 3, 954-961. (2003)
15. Robert A., Katz, Stephen M., Pizer: Untangling the Blum Medial Axis Transform. *International Journal of Computer Vision*, vol. 55, No. 2-3, 139-153. (2003)

Xiaopeng SUN, J. PAN, and Xiaopeng WEI

16. Ma, W., Wu, F., Ouhyoung M.: Skeleton Extraction of 3D Objects with Radial Basis Functions. IEEE Proceedings of Shape Modeling International. Seoul, Korea, 207-215. (2003)
17. Telea, A., Vilanova A.: A robust level-set algorithm for centerline extraction, in Eurographics/IEEE Symposium on Visualization, Grenoble, France , 185–194. (2003)
18. Zhang, X., Liu, J., Jaeger Z. Li. M.: Volume decomposition and hierarchical skeletonization. In VRCAI '08: Proceedings of The 7th ACM International Conference on Virtual-Reality Continuum and Its Applications in Industry. Singapore; December. (2008)
19. Bertrand, G.: A parallel thinning algorithm for medial surfaces. Pattern Recognition Letters, vol. 16, No. 9, 979-986.(1995)
20. Osher, S., Sethian J.: Fronts propagating with curvature dependent speed: algorithm based on the Hamilton-Jacobi formulation, Journal of Computational Physics, vol 79, 1, 12-49. (1988)
21. Hoffman, D.D., Richards, W.A.: Parts of recognition. Cognition, vol. 18, 65-96. (1985)
22. Rom, H., Medioni, G.: Hierarchical decomposition and axial shape description. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, No. 10, 973-981. (1993)
23. Xiaobai Chen, Aleksey Golovinskiy, Thomas Funkhouser A.: A benchmark for 3D mesh segmentation. ACM Trans. Graph. Vol. 28, No. 3. (2009)
24. Lien, J. M., Keyser, J., Amato, N. M.: Simultaneous shape decomposition and skeletonization. In Proceedings of the 2006 ACM symposium on Solid and physical modeling, Cardiff, Wales, UK, 219-228. (2006)
25. Sagi Katz, George Leifman, Ayellet Tal: Mesh segmentation using feature point and core extraction, The Visual Computer, vol. 21, No. 8/10, 649-658. (2005)
26. Vitaly Surazhsky, Tatiana Surazhsky, Danil Kirsanov, Steven J. Gorthier, Hugues Hoppe: Fast Exact and Approximate Geodesics on Meshes. ACM Trans. on Graphics, vol. 24, No. 3, 553-560. (2005)
27. Verbeek, J. J., N, Kr se B Vlassis: A soft K-segments algorithm for principal curves. In: Proceedings of International Conference on Artificial Neural Networks, Vienna, 450-456. (2001)
28. Michael Elad, Ayellet, Sigal Ar: Content Based Retrieval of VRML Objects-An Iterative and Interactive Approach. In: Proceedings of the sixth Eurographics Workshop on Multimedia, Manchester UK, 107-118.(2001)

Xiaopeng Sun was born in P. R. China and received a PhD in Computer Science from Institute of Computing Technology, Chinese Academy of Sciences. He is a distinguished professor of LNNU, and his research interests include graphics and VR.

Xiaopeng Wei was born in P. R. China and received a PhD in Mechanical Engineering from Dalian University of Technology. He is a professor at DLUT, and his research interests include graphics.

Received: April 29, 2009; Accepted: October 03, 2009.

A Dynamic Alignment Algorithm for Imperfect Speech and Transcript

Ye Tao^{1,2}, Xueqing Li¹, and Bian Wu²

¹Shandong University, Department of Computer Science and Technology,
Ji Nan, 250101, P. R. China

²Shanghai Qitai Internet Technology Co. Ltd., Shanghai, 201203, P. R. China

Abstract. This paper presents a novel alignment approach for imperfect speech and the corresponding transcription. The algorithm gets started with multi-stage sentence boundary detection in audio, followed by a dynamic programming based search, to find the optimal alignment and detect the mismatches at sentence level. Experiments show promising performance, compared to the traditional forced alignment approach. The proposed algorithm has already been applied in preparing multimedia content for an online English training platform.

Keywords: Text-Audio Alignment; Dynamic Programming

1. Introduction

The motivation of this research comes from a content producing module of an English tutor platform [1], which tests and evaluates learners' spoken English level as a foreign language. To perform pronunciation analysis, it is necessary to have the time-aligned word/phoneme transcriptions with audio data. In fact, time-aligned labels can be used not only in the area of language training [11] [3], they are also required by audio/video indexing techniques applied in search engines [15] [4] [5]. Moreover, as a fundamental task in speech processing, it could be useful in model training for Automatic Speech Recognition (ASR) systems [21]. Though aligning speech with its corresponding text might seem a solved problem [8], situation could be difficult if the transcription does not match the media, as the decoder is forced to accept the input transcription. Unfortunately, audio files and their transcriptions are not fully-matched in many cases. Alignment of audio and text with imperfections has applications in subtitling, spoken books and etc. For example, a dialogue script often includes speaker names which are not uttered in audio to make it more readable, and broadcast/TV news scripts sometimes skip speech from interviewees. Manually scanning the discrepancies is a tedious and time-consuming work that requires skill.

Recently, alternate approaches have been investigated to align the speech with its approximate text. Moreno et al [10] developed a recursive method to progressively reduce the forced alignment process with a gradually restricting

dictionary and language model. A similar approach has been presented in [7] for human generated transcriptions of audio files. These techniques are based on the methodology of comparing ASR result with the approximate transcript. However, the recognition quality highly depends on the acoustic model, and thus could be severely degraded without speaker adaption and appropriate model training. In contrast, most Viterbi-based forced alignment algorithms give satisfactory result, even the acoustic environment of the input speech is quite different from the one used for model training. Another limitation of Moreno's approach is that the anchor selector is difficult to handle repeated words, phrases and sentences, which are widely used in language tutor. In this case, some recognized anchors can become ambiguous in the original text, and cause alignment errors. Experiment in section 4 indicates that the performance of this approach is relatively low on teaching materials and online courses, even for perfect matched audio and text.

Other authors have included HMM garbage models to allow for text/speech skips and substitutions [12], which works well in discovering and correcting low level (word/phrase) errors, where most parts are matched. The approach presented in this paper focuses more on detecting the insertion and deletion errors at sentence/paragraph level. We convert the alignment problem into a series of overlapping sub-problems, which are solved recursively by a dynamic programming algorithm. The algorithm has been implemented in a content producing system, processing speeches from varied sources, such as news, online courses, lectures and etc.

The rest of the paper is organized as follows. Section 2 introduces the methodology of sentence boundary detection. Section 3 focuses on the dynamic alignment algorithm and the pruning policy. Section 4 gives some preliminary experiment results and section 5 concludes the paper.

2. Pre-processing

Input audio and text are required to be segmented into small (e.g. sentence) unit. Transcript can be segmented by using the maximum entropy approach [2], which is one of the state-of-the-art natural language processing techniques. Speech sentence boundary detection is much more challenging, since typical cues in text (e.g. headers, paragraphs, punctuation and etc.) are absent in utterances [17]. Quite a few jobs have been done in automatic detection of prosodic boundaries in speech [20] [16]. We use a multi-stage pre-processing approach to find the approximate sentential boundaries, as shown in Fig. 1.

1. The source audio file provided by user often contains non-speech parts, e.g. lectures with prelude and epilogue music are very common in multimedia courses. Firstly, these non-speech clips are separated and removed. Different strategies have been investigated for speech/non-speech detection during the last decade [9]. And our previous research [19]

also proposed a fuzzy logic based approach that combines different features to label the boundaries of voice segments.

2. Pauses detection on speech can typically be done with a high accuracy off-line VAD algorithm. Numerous solutions have been reported to achieve precise detection results [13]. The method we used in this system is based on the Order Statistics Filtering Sub-Band Spectral Entropy [6], which measures the sub-band spectrum divergence between speech and background noise. Long-term speech features [14] can also be considered as contextual information to estimate the threshold more precisely and benefits for detecting speech presence in noisy environments.
3. Finally, boundaries detected by the VAD algorithm need to be filtered to get the prosodic boundaries. It has been studied in previous research that a sentence boundary is often marked by some combination of a long pause, a preceding final low boundary tone, and a pitch range reset [18]. Therefore, pause information can be important cues to eliminate inner-sentence boundaries.

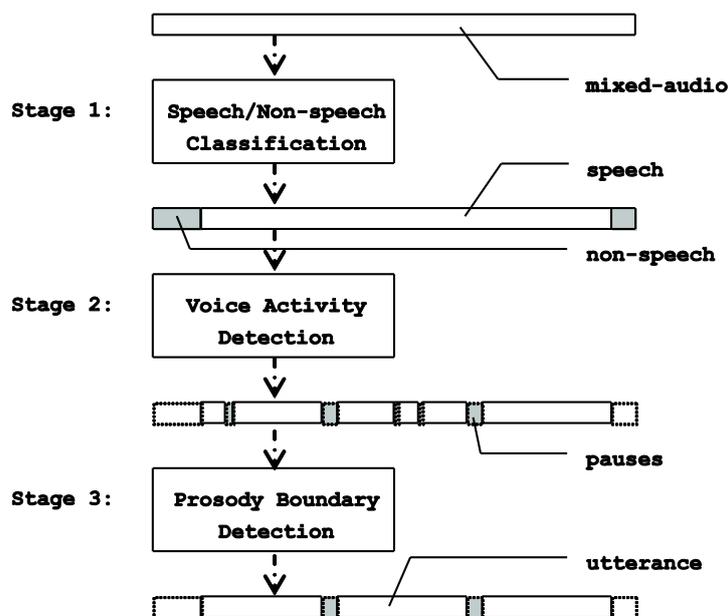


Fig. 1. 3-stage sentence boundary detection

Sentence segmentation is a very difficult problem in spontaneous speech such as lectures, and thus addressed by many works recently. It is easily possible that sentential boundary detection based on prosodic cues can produce errors causing split or concatenated sentences within this pre-processing stage. An algorithm designed for correcting these errors (including false alarms and missed alarms) is presented in the following section.

3. Dynamic Alignment

Let $s(i,j)$ and $t(m,n)$ denote the utterance and transcript whose boundaries are positioned at i, j and m, n , where $i \leq j$ and $m \leq n$. We are going to find the most feasible sentential matches between the text and audio, i.e.

$$\arg \max_{1 \leq i \leq j \leq N, 1 \leq m \leq n \leq M} s(i, j) \leftrightarrow t(m, n), \quad (1)$$

where M and N are the number of segments in speech and text, $s(i,j) \leftrightarrow t(m,n)$ measures the likelihood of the alignment.

3.1. Dynamic programming

If the audio and text are presented in order, the problem can be broken up into stages with an alignment required at each stage. Let the ordered pair (h,k) denotes a hypothetic alignment for $s(0,h) \leftrightarrow t(0,k)$. To find the best solution at stage (h,k) , it is necessary to go through all the possible matches in previous stages, and see how to make an alignment for the remainder. Denoted by $F(h,k)$ as the maximum similarity accumulation at stage (h,k) , we have the following induction:

$$F(h, k) = \max_{t_h, t_k} F(h - t_h, k - t_k) + P(h - t_h, h, k - t_k, k), \quad (2)$$

$$(0 < h < N, 0 < k < M, t_h \in [0, \varepsilon], t_k \in [0, \varepsilon], t_h + t_k \neq 0)$$

where ε is the width of search beam. $P(i,j,m,n)$ computes the acoustic likelihood of the alignment for $s(i,j) \leftrightarrow t(m,n)$, which indicates the strength of belief that how much they are matched.

3.2. Alignment function

Fig. 2 presents the alignment result of an utterance and its corresponding transcript, where the value of P is computed as follows:

$$P(i, j, m, n) = \left\{ \sum_{[f_i, f_{bs}] \cup [f_{be}, f_j]} \omega_b \times Acc + \sum_{[f_{bs}, f_{be}]} \omega_c \times Acc \right\} / (f_j - f_i), \quad (3)$$

$$(i < j, m < n)$$

where f_i and f_j are the start/end frame indices of the utterance $s(i,j)$, f_{bs} and f_{be} define the boundaries frame indices, ω_b and ω_c specify the weights of the boundaries and internal parts respectively, and Acc is the normalized acoustic score.

Acoustic score value indicates the likelihood that a speech segment represents a particular symbol according to the statistical models. However, the value depends on the length of the segment, and thus needs to be normalized for a particular segment, which simply entails dividing the score by the number of frames contained in the segment. *Acc* thus represents the average log likelihood per frame for the given segment, and can be used to compare speech segments of different lengths (typically different phones) to determine which segments fit better. In addition, *P* needs to be distinguishable between the fully-matched (e.g. $s(1,2) \leftrightarrow t(0,1)$) and partially-matched pairs (e.g. $s(0,2) \leftrightarrow t(0,1)$), and thus weights are at boundaries to guide the solver towards the global optimal solution.

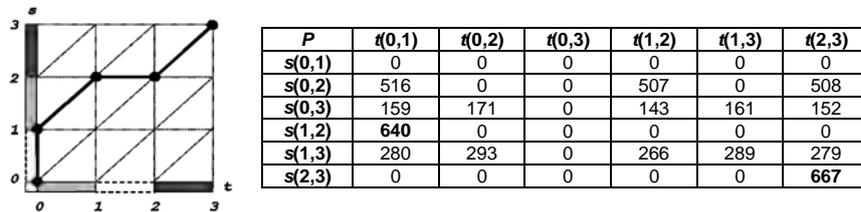


Fig. 2. An example of dynamic alignment. $s(0,1)$ is unuttered and $t(1,2)$ is untranscribed

3.3. Sentential Boundary Correction

As discussed earlier, speeches (e.s.p. unprepared speech or conversational speech) often contains pauses due to speech errors, false starts, train-of-thought gaps and etc, which causes false and missed alarms in sentential boundaries detection in pre-processing stage. However, it is possible to correct most of these errors by scanning and comparing the forced alignment results.

False alarm

We compare the alignment results of successive speeches upon the same text (w_1, \dots, w_N), i.e. from $P(i, i+1, m, n)$ to $P(i, i+L, m, n)$, where L is a pre-defined value to control the size of the search window, as shown in Fig. 3(a). When a monotonically increasing is detected, i.e.

$$P(i, i+1, m, n) < \dots < P(i, i+L, m, n) \tag{4}$$

which indicates that alignments are stably improved by extending the speech, we then remove the last L_t inner boundaries, where $0 \leq L_t < L$ is a threshold.

Missed alarm

Fig. 3(b) shows an example of the missed alarm detection. Our forced alignment engine uses a generic speech model (garbage model) to absorb the out-of-vocabulary words, thus the notable detected silence at the end of the utterance indicates the existence of a text-skip, and in this case, a new sentence boundary should be inserted at the end of word w_N .

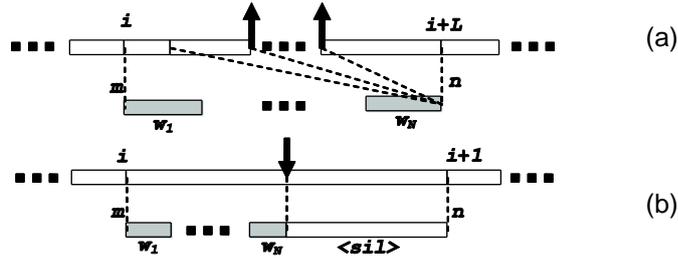


Fig. 3. Sentential boundary correction. (a) false alarm. (b) missed alarm

3.4. Pruning strategy

Pruning strategies are applied to eliminate most unlikely search hypotheses. A letter-to-sound algorithm can be used to predict the length of a sentence, by simply counting the number of words and phones. This could be helpful to avoid attempts on most impossible alignments. If continuous confidence score descents are detected on two utterances upon the same text, e.g. $P(1,2,0,1)$ and $P(1,3,0,1)$ in Fig. 2, we then eliminate any successor, by setting P to 0. Due to the antithesis of this problem, another pruning rule can be applied in text domain, as stated below, where γ^* are pre-defined thresholds.

$$\begin{aligned}
 P(i, j, m, n) - P(i, j-1, m, n) &\leq \gamma_1 \rightarrow P(i, x, m, n) = 0, \\
 P(i, j, m, n) - P(i, j, m, n-1) &\leq \gamma_2 \rightarrow P(i, j, m, y) = 0, \\
 (0 < i < j < N, 0 < m < n < M, x \in [j+1, M], y \in [n+1, N]).
 \end{aligned} \tag{5}$$

Moreover, to accelerate the solving speed, we save those P and F we have already computed. If we need to solve the same problem later, we then retrieve and reuse our already-computed values.

4. Experiment and Discussion

Alignment performance was evaluated on a data set collected from lectures (15%, speeches and interviews), multimedia courses (65%, English teaching

materials for K-12 students), broadcasting/television news (20%, live news and BBC/VOA special programs). Table. 1 summarizes the experimental data set, where each type of data is a composition of clips that are fully-matched (CP), and clips that contains mismatched parts (CI), to test the robustness and compatibility of the proposed algorithm.

Table 1. Data set and comparison experiment results.

types	clips	sentences P/I	length (min.)	F.A	A.A	D.A	
						$\epsilon = 1$	$\epsilon = 2$
lectures	P	37/-	7	37(100%)	30(81%)	33(90%)	34(92%)
	I	38/14	9	33(63%)	39(75%)	45(86%)	48(92%)
news	P	32/-	10	32(100%)	28(88%)	26(81%)	28(88%)
	I	39/16	11	28(50%)	39(71%)	47(74%)	47(74%)
courses	P	80/-	29	80(100%)	61(76%)	69(86%)	73(91%)
	I	49/44	37	54(58%)	69(74%)	73(78%)	77(83%)

Fig. 4 shows the sentence boundary detection results, where false alerts were the major source of errors for the pure VAD algorithm (stage 1). Richness of such errors is related to the corpus we chose, as many of the speeches are designed for teaching and thus contains long inner-sentence pauses. The existence of missed boundary error often related to the variability in the user speaking state, e.s.p. when the user tends to speed up the speech at the end of a sentence. Pause (stage 2) and pitch (stage 3) information are helpful to reduce the false alarm rate. And the VAD correction algorithm (Section 3.3) also provides a sustained improvement in both sentence boundary hit rate and false alarm rate over the 3-stage sentence boundary detection algorithm.

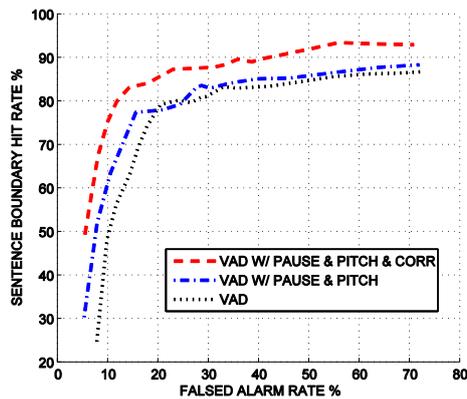


Fig. 4. Results of sentence boundary detection

The acoustic score in P is achieved with a basic speaker independent recognizer tuned to run in the forced alignment mode. An alignment of a

speech-text pair $s(i,j) \leftrightarrow t(m,n)$ is defined as successful, when the normalized confidence score of exceeds a threshold, i.e. $P(i,j,m,n) > T$, where T is a pre-defined value to guarantee that there is no alignment flaws.

As shown in Table. 1, we found that the anchor-based algorithm (A.A) may fail to give the expected result for fully-matched test cases, when the news and lectures are record in a noisy environment. In particular, the performance of A.A. degrades seriously on those clips designed for teaching and learning, due to a large amount of repeat words and phrases. For speech and text that are not well-matched, our approach significantly increases the ratio of successful alignment, compared to the traditional forced alignment.

An examination of the results shows that most failures are caused by the consecutive mismatches. Performance highly depends on the quality of the sources, e.s.p. the ratio of the mismatched parts, and ϵ a trade-off between accuracy and speed. Better results can be achieved by setting a wider search beam, e.g. changing ϵ from 1 to 2 will increase the correct alignment ratio, it will however direct the algorithm to try more possibilities and slow down the alignment process. In general, on each stage, the number of search paths C and width of search beam follows:

$$C = \epsilon^2 + 2\epsilon \tag{6}$$

Pruning policies limits the range of alignments and removes most infeasible searches, and it can reduce the amount of computation, as shown in Fig 5.

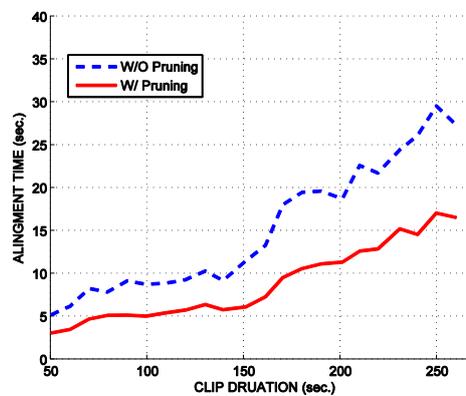


Fig. 5. Results of pruning. $\epsilon = 1$, evaluated on a P4 2.4GHz computer with 2G RAM installed

5. Conclusion

We introduced an approach for the temporal alignment of speech with imperfect transcripts, based on the acoustic likelihood marked chunks of

speech signals, that are associated with partitioned audio segments, and a word level symbol sequence given in the erroneous transcription. In this paper, speech and text are first segmented into units, which are then aligned with a dynamic alignment algorithm. The proposed algorithm has been implemented and validated by an easy-to-use content producing tool for preparing multimedia content for English training. The experiment result shows an increase of the correct matching ratio, in particular for those clips whose speech and transcription are not well-matched, compared to the traditional forced alignment approach.

Though most speech and text are presented in order, a limitation of the algorithm is it is not efficient for the re-ordering of phrases or sentences in transcription. Future investigations also include launching ASR on only the mismatched parts, using the idea described in [10] in conjunction to correct the errors and accelerate the detection speed.

6. References

1. <http://www.mytalkpal.com>.
2. Adam L. Berger, Vincent J. Della Pietra, and Stephen A. Della Pietra. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1), 39-71. (1996)
3. Jared Bernstein, Michael Cohen, Hy Murveit, Dimitry Rtischev, and MitchelWeintraub. Automatic evaluation and training in english pronunciation. In *Proc. of ICSLP 90*, 1185-1188. (1990)
4. Satya Dharanipragada, Martin Franz, and Salim Roukos. Audio-indexing for broadcast news. In *Proc. of TREC6*, 115-119. (1997)
5. Jonathan Foote. An overview of audio information retrieval. *ACM Multimedia Systems*, 7, 2-10. (1999)
6. Zhang Ya-Xin Lv Yue Guo Li-Hui, He Xin. An order statistics ltering-based real-time voice activity detection algorithm. *ACTA AUTOMATICA SINICA*, 34(4), 419-425. (2008)
7. Timothy J. Hazen. Automatic alignment and error correction of human generated transcripts for long speech recordings. In *Proc. of ICSLP 06*, 1606-1609. (2006)
8. Picone J., Goudie-Marshall K., Doddington G., and Fisher W. Automatic text alignment for speech system evaluation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(4), 780-784. (1986)
9. Lie Lu, Hong-Jiang Zhang, and Hao Jiang. Content analysis for audio classification and segmentation. *IEEE Transactions on Speech and Audio Processing*, 10(7), 504-516. (2002)
10. P. Moreno, C. Joerg, J.-M. Van Thong, and O. Glickman. A recursive algorithm for the forced alignment of very long audio segments. In *Proc. of ICSLP 98*, 2711-2714. (1998)
11. Leonardo Neumeyer, Horacio Franco, Mitchel Weintraub, and Patti Price. Automatic text-independent pronunciation scoring of foreign language student speech. In *Proc. of ICSLP 96*, 1457-1460. (1996)
12. Long Nguyen and Bing Xiang. Light supervision in acoustic model training. In *Proc. of ICASSP 04*, 1, 185-188. (2004)
13. J. Ramirez, J. M. Gorriz, and J. C. Segura. *Robust Speech Recognition and Understanding, Chapter 1: Voice Activity Detection. Fundamentals and Speech*

Ye Tao, Xueqing Li, and Bian Wu

- Recognition System Robustness, 460-481. I-Tech Education and Publishing. (2007)
14. J. Ramirez, J. C. Segura, Carmen Benitez, Angel de la Torre, and Antonio Rubio. Efficient voice activity detection algorithms using long-term speech information. *Speech Communication*, 42(3-4), 271-287. (2004)
 15. Deb Roy and Carl Malamud. Speaker identification based text to audio alignment for an audio retrieval system. In Proc. of ICASSP 97, 1099-1102. (1997)
 16. Elizabeth Shriberg, Andreas Stolcke, Dilek Hakkani-Tur, and Gukhan Tur. Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication*, 32(1-2), 127-154. (2000)
 17. Mark Stevenson and Robert Gaizauskas. Experiments on sentence boundary detection. In Proceedings of the sixth conference on Applied natural language processing, 84-89. (2000)
 18. Marc Swerts and Mari Ostendorf. Prosodic and lexical indications of discourse structure in human-machine interactions. *Speech Communication*, 22(1), 25-41. (1997)
 19. Ye Tao, Daren Zu, Xueqing Li, and Ping Du. A fuzzy logic based speech extraction approach for e-learning content production. In Proc. of ICALIP 2008, 1, 298-302. (2008)
 20. Arthur R. Toth. Forced alignment for speech synthesis databases using duration and prosodic phrase breaks, Fifth ISCA Workshop on Speech Synthesis, (2004)
 21. Steve Young, Gunnar Evermann, Thomas Hain, Dan Kershaw, Gareth Moore, Julian Odell, Dave Ollason, Dan Povey, Valtcho Valtchev, and Phil Woodland. The HTK Book (for HTK Version 3.2.1). Cambridge University Engineering Department, (2002)

Ye Tao is working for his PhD in Department of Computer Science and Technology, Shandong University, P.R.China. He obtained M. Tech. Degree in software engineering from Shandong University. His areas of interests are Speech Processing and Computer Graphics.

Xueqing Li is a Professor in Department of Computer Science and Technology, Shandong University, P.R.China. He obtained BSc, MSc, and PhD all from Shandong University. His research interests include Human Computer Interaction & Virtual Reality, Computer Graphics, Image Processing and Computer Geometry.

Bian Wu is the CTO of Shanghai Qitai Internet Technology Co. Ltd., P.R.China. He obtained PhD from Shanghai Jiaotong University. His research interest is Speech Processing.

Received: July 25, 2009; Accepted: October 28, 2009.

Multi-Video Summarization Using Complex Graph Clustering and Mining

Jian Shao¹, Dongming Jiang¹, Mengru Wang², Hong Chen²,
and Lu Yao¹

¹ College of Computer Science, Zhejiang University, China

² Zhejiang Radio & TV Group, China

jshao@cs.zju.edu.cn, dmjiang1985@163.com, cs08yl@hotmail.com
{wmr628, ch213}@mail.zrtg.com

Abstract. Multi-video summarization is a great theoretical and technical challenge due to the wider diversity of topics in multi-video than single-video as well as the multi-modality nature of multi-video over multi-document. In this paper, we propose an approach to analyze both visual and textual features across a set of videos and to create a so-called circular storyboard composed of topic-representative keyframes and keywords. We formulate the generation of circular storyboard as a problem of complex graph clustering and mining, in which each separated shot from visual data and each extracted keyword from speech transcripts are first structured into a complex graph and grouped into clusters; hidden topics in the representative keyframes and keywords are then mined from clustered complex graph while at the same time maximizing the coverage of the summary over the original video set. We also design experiments to evaluate the effectiveness of our approach and the proposed approach shows a better performance than two other storyboard baselines.

Keywords: multi-video summarization, complex graph clustering and mining, circular storyboard.

1. Introduction

Multi-video summarization aims to create a summary for a set of topic-related videos having several sub-themes or sub-events around a main topic. Imagine a scenario that a news aggregator gathers various sets of topic-related news videos and provides summaries to represent each video set, such that the users can easily understand the topics after taking a look at the summaries and decide whether to go through watching their interested video set.

Despite that a lot of approaches of automatic video summarization [1-5] and some of multi-document summarization [6] have been proposed in literature, little work has focused on multi-video [7]. In general, multi-video summarization faces two difficult challenges. One is due to the inevitable

thematic diversity and overlaps within multi-video than single-video, and hence we need study effective summarization methods to extract the globally main topic information while removing redundancy among different videos as more as possible. The other is due to the multi-modality nature of multi-video over multi-document. Using both text and imagery during summarization is more effective than either modality alone [5, 8], and hence we need effective summarization methods to integrate both textual and visual content in summary creation and visualization.

An intuitive approach for summarization of multi-video with text transcripts is to summarize visual modality and textual modality separately and to visualize as an image-plus-text list storyboard together — employ a shot clustering and mining algorithm to select a list of most representative keyframes from shot clusters to reflect the visual topics, employ multi-document summarization approaches to select a list of most representative keywords to reflect the textual topics, generate a list storyboard by placing a list of keyframes in upper half and a list of keywords in lower half. However, despite that this list storyboard gives brief summaries of both visual content and textual content, it hardly makes use of the relations of visual content and textual content. Actually, given a video set, there usually exist a number of sub-themes, and each sub-theme can be represented as a collection of shots and keywords which have thematic relations with each other. Discovering the relations between shots and keywords will not only offer benefits to find more reasonable topic structure, but also make it possible to further remove redundancy in theme level.

To take advantage of both visual and textual information during summary generation, we propose a novel approach to summarize visual modality and textual modality simultaneously and to visualize as an image-plus-text circular storyboard in this paper. We first build a complex graph consisting of shot nodes and keyword nodes. Shots are linked to each other by visual similarity, while shots and keywords are linked to each other by co-occurrences. In such a way, rather than one-way clustering, i.e., either shot clustering or keyword clustering, we can perform co-clustering of shots and keywords on complex graph to make use of the benefits of shot-keyword relations [9]. Next, we mine representative keyframes and keywords for hidden topics from clustered complex graph with highest importance score while at the same time maximizing the coverage of the summary over the original video set by deeply exploiting the node-level relations and the cluster-level relations. And finally, we design a circular storyboard to present the visual and textual topics and their relations — resize the keyframes and keywords according to their importance and arrange resized keyframes and keywords around or inside a circle according to their relation strength. Experiments carried out on video news demonstrate the effectiveness of our proposed approach.

We summarize our main contributions as follows: 1) The novel utilization of complex graph clustering for multi-video summarization; 2) The scheme of mining representative keyframes and keywords for hidden topics by integrating both node-level and cluster-level information in clustered complex

graph; 3) the visual presentation of a summary as an image-plus-text circular storyboard.

The remainder of this paper is organized as follows. Section 2 describes our complex graph clustering and mining based multi-video summarization approach. Section 3 describes the data set and experimental results. Finally, we conclude this paper in Section 4.

2. The Multi-Video Summarization Approach by Complex Graph Clustering and Mining

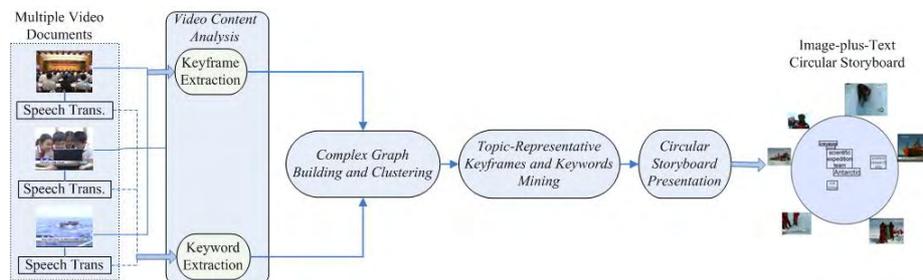


Fig. 1. Proposed multi-video summarization approach

Our proposed multi-video summarization approach is shown in Fig. 1, where the input is multiple video documents with their corresponding speech transcripts, and the output is so-called circular storyboard, an image-plus-text static summary composed of representative keyframes and keywords of hidden topics. There are four stages to generate the summary, of which the first one is to analyze visual content to extract a pool of shots and to analyze textual content of speech transcripts to extract a list of keywords across multiple video documents; the second is to perform complex graph building and clustering on extracted shots and keywords; the third is to mine the representative keyframes and keywords of hidden topics from clustered complex graph; and the last is to use a circle space to present the topic-representative keyframes and keywords as circular storyboard.

2.1. Video Content Analysis

Assume that we have a set of video documents with their speech transcripts, $D = \{d_1, \dots, d_m, \dots, d_M\}$. To analyze the visual content of given videos, we employ a robust shot boundary detection algorithm in [10] to divide video sequences into a pool of shots denoted as $U = \{u_1, \dots, u_i, \dots, u_I\}$, where u_i is a shot and I denotes the number of shots in video set. For further visual content processing, we select the middle frame of a shot as the keyframe and represent it as a 423-dimensional vector of 6 features: 256-dim color

histogram, 6-dim color moments, 128-dim color coherence, 15-dim texture MSRSAR, 10-dim texture Tamura coarseness, and 8-dim texture Tamura directionality.

In addition, in order to analyze the textual content of given videos, a three-step process is conducted to extract meaningful keywords from speech transcripts. First, a part of speech tagger is used to label out nouns in speech transcript. Next, stop-words are used to further filter unnecessary nouns. And finally, we propose a modified tf-idf formula to score the remaining keywords and select the keywords with highest importance score. Suppose that the selected keywords are denoted as $W = \{w_1, \dots, w_j, \dots, w_J\}$, the importance score $I(w_j)$ of keyword w_j based on the modified tf-idf formula is given by

$$I(w_j) = \text{mod_tf}(w_j, D) \times \text{idf}(w_j) \quad (1)$$

where $\text{idf}(w_j)$ represents the inverse document frequency and can be defined as $\text{idf}(w_j) = \log(N / df(w_j))$, where N is the total number of videos in training corpus, and $df(w_j)$ is the number of videos involving the keyword of w_j . $\text{mod_tf}(w_j, D)$ represents the modified term frequency of word w_j in video set D , it is defined by

$$\begin{aligned} \text{mod_tf}(w_j, D) &= \sum_{m=1}^M \text{tf}(w_j, d_m) \times \left(\frac{\text{tf}(w_j, d_m)}{\text{tf}(w_j, D)} + \alpha \times \left(1 - \frac{\text{tf}(w_j, d_m)}{\text{tf}(w_j, D)} \right) \right) \\ \text{tf}(w_j, D) &= \sum_{m=1}^M \text{tf}(w_j, d_m) \end{aligned} \quad (2)$$

where $\text{tf}(w_j, d_m)$, $\text{tf}(w_j, D)$ represents the frequencies of word w_j in video d_m , in video set D respectively, α is weighting factor. Note that in contrast with $\text{tf}(w_j, D)$, $\text{mod_tf}(w_j, D)$ takes the keyword distribution of video set into account. By setting $\alpha > 1$, the more disperse the keyword distribution of video set is, the larger the $\text{mod_tf}(w_j, D)$ is, and correspondingly the more important the keyword is.

2.2. Complex Graph Building and Clustering

Considering the task of learning cluster structure from a pool of shots U and a list of keywords W that are extracted from video set D , we can use most one-way existing clustering algorithm to cluster shots and keywords separately [11], or employ co-clustering algorithm [9, 12] to cluster shots and keywords simultaneously. In this paper, we adopt the complex graph clustering algorithm in [9] to simultaneously derive the shot clusters and keyword clusters as well as their relations.

We first organize extracted shots U and keywords W to build a complex graph of shot nodes and keyword nodes. Suppose the instantiated complex graph is denoted as $G=(V^{(1)},V^{(2)},E^{(1,1)},E^{(1,2)})$, where $V^{(1)}=\{v_1^{(1)},\dots,v_i^{(1)},\dots,v_I^{(1)}\}$ represents the shot node set, $v_i^{(1)}$ is the i th node in $V^{(1)}$ corresponding to the i th shot in U ; $V^{(2)}=\{v_1^{(2)},\dots,v_j^{(2)},\dots,v_J^{(2)}\}$ represents the keyword node set, $v_j^{(2)}$ is the j th node in $V^{(2)}$ corresponding to the j th keyword in W ; $E^{(1,1)}$ represents the homogeneous edges within shot nodes; $E^{(1,2)}$ represents the heterogeneous edges between shot nodes and keyword nodes. We use affinity matrix $S \in R_+^{I \times I}$ to represent the weights of $E^{(1,1)}$ and use $A \in R_+^{I \times J}$ to represent the weights of $E^{(1,2)}$. Cosine similarity metric is employed to realize matrix S . That is, the edge weight $S_{i,q}$ between node $v_i^{(1)}$ and node $v_q^{(1)}$ is given by

$$S_{i,q} = \frac{fe(u_i).fe(u_q)}{|fe(u_i)| \times |fe(u_q)|} \tag{3}$$

where $fe(u_i)$ denotes the keyframe feature vector of shot u_i . In addition, The co-occurrence between shots and keywords is employed to realize matrix A . That is, the edge weight $A_{i,j}$ between shot node $v_i^{(1)}$ and keyword node $v_j^{(2)}$ is given by

$$A_{i,j} = \sum_{k=1}^{tf(w_j)} sim_t(t(u_i),t(w_j^k)) \tag{4}$$

Note that word w_j may occur several times $tf(w_j)$ in a video which contains shot u_i , so every occurrence is indexed by k . Timing similarity $sim_t(t(u_i),t(w_j^k))$, between $t(u_i)$, the mid-point timing of shot u_i , and $t(w_j^k)$, the mid-point timing of the k th occurrence of word w_j , is defined as follows:

$$sim_t(u_i, w_j^k) = \begin{cases} \exp\{-\frac{|t(u_i)^{start} - t(w_j^k)|^2}{2\sigma_t^2}\} & (t(w_j^k) < t(u_i)^{start}) \\ 1 & (t(u_i)^{start} \leq t(w_j^k) \leq t(u_i)^{end}) \\ \exp\{-\frac{|t(w_j^k) - t(u_i)^{end}|^2}{2\sigma_t^2}\} & (t(w_j^k) > t(u_i)^{end}) \end{cases} \tag{5}$$

This $sim_t(.,.)$ function is a step function having 1 when $t(w_j^k)$ falls in the range between $t(u_i)^{start}$, the start-point timing of shot u_i , and $t(u_i)^{end}$, the end-point timing of shot u_i , but its edges are dispersed using a Gaussian filter with standard deviation σ_t to compensate for the time delay between the shot and keyword occurrence.

We then perform a hard clustering algorithm described in [9] to derive the shot clusters and keyword clusters as well as their relations between clusters. Suppose that we have grouped the shot nodes $V^{(1)}$ into K clusters $CU = \{cu^{(1)}, \dots, cu^{(k)}, \dots, cu^{(K)}\}$ and grouped the keyword nodes $V^{(2)}$ into L clusters $CW = \{cw_1, \dots, cw_l, \dots, cw_L\}$. Let $C^{(1)} \in (0,1)^{I \times K}$ denote the cluster membership matrix for shot nodes $V^{(1)}$ such that $C_{i,k}^{(1)}$ denote the weight that the i th shot node in $V^{(1)}$ is associated with the k th cluster, and $C^{(2)} \in (0,1)^{J \times L}$ denotes the cluster membership matrix for keyword nodes $V^{(2)}$ such that $C_{j,l}^{(2)}$ denote the weight that the j th keyword node in $V^{(2)}$ is associated with the l th cluster. The intra-type cluster relation matrix $D \in R^{K \times K}$ denotes cluster relations within the same type of nodes such that $D_{k,r}$ denotes the link strength between the k th shot cluster cu_k and the r th shot cluster cu_r . The inter-type relation matrix $B \in R^{K \times L}$ denotes the cluster relations between the different types of nodes such that $B_{k,l}$ denotes the link strength between the k th shot cluster cu_k and the l th keyword cluster cw_l .

2.3. Topic-Representative Keyframes and Keywords Mining

This step aims to mine the most representative keyframes and keywords of hidden topics from clustered complex graph. We propose a three-stage mining scheme.

First of all, we define measures to compute the importance of shot nodes and keyword nodes. We employ the modified tf-idf formula described in Sub-Section 2.1 to calculate the importance of keyword nodes. While for computing the importance of shot nodes, we take both the visual and related textual features into consideration. Given a shot u_i in node $v_i^{(1)}$, its importance score $I(u_i)$ is modeled as a combination of its visual informativeness $\text{inf}_{\text{visual}}(u_i)$ and related textual informativeness $\text{inf}_{\text{text}}(u_i)$ with a weight parameter β specified by users:

$$I(u_i) = \beta \times \text{inf}_{\text{visual}}(u_i) + (1 - \beta) \times \text{inf}_{\text{text}}(u_i) \quad (6)$$

The computation of textual informativeness of a shot is based on the following intuitions:

- the more important a shot's related keywords are, the more informative it is;
- the more heavily a shot is linked with related words, the more informative it is.

Based on above heuristic intuitions, $\text{inf}_{\text{text}}(u_i)$ can be expressed by the following formula:

$$\text{inf}_{\text{text}}(u_i) = \sum_{j=1}^J I(w_j) \times A_{i,j} \quad (7)$$

On the other hand, the computation of the visual informativeness of a shot is based on the following intuitions:

- the more a shot's similar shots are, the more informative it is;
- the more a shot's similar shots in other videos are, the more informative it is.

Given above heuristic intuitions, $\text{inf}_{\text{visual}}(u_i)$ can be expressed by the following formula:

$$\text{inf}_{\text{visual}}(u_i) = \sum_{i \neq p} S_{i,p} \times \delta(u_i, u_p, D) \quad (8)$$

$$\delta(u_i, u_j, D) = \begin{cases} 1 & \text{if } D(u_i) = D(u_j) \\ \gamma & \text{otherwise} \end{cases}$$

where $\delta(u_i, u_p, D)$ is a weight function, and we set $\gamma > 1$ to differentiate the shot edges within one video or across two videos.

Next, we define measures to compute the informativeness of shot clusters and keyword clusters. Different from single-video, multi-video usually contains several sub-themes or sub-events and each sub-theme of sub-event can be viewed by a cluster of theme-related keywords or a cluster of event-related shots. Thus, the computation of the importance of a keyword cluster is based on the following intuitions:

- The more complex a cluster is, the more important it is
- The more important a cluster's contained keywords are, the more important it is.

Based on above intuitions, $I(cw_l)$, the importance of l_{th} keyword cluster cw_l is given by

$$I(cw_l) = \sum_{j=1}^J C_{j,l}^{(2)} \times I(w_j) \quad (9)$$

While for computing the importance score of a shot cluster, we take both the shot nodes inside cluster and the relations with other clusters into consideration. Given a shot cluster cu_k , its importance score $I(cu_k)$ is modeled as a combination of its node-level informativeness $\text{inf}_{\text{node}}(cu_k)$ and cluster-level informativeness $\text{inf}_{\text{clust}}(cu_k)$ with a weighting factor ξ :

$$I(cu_k) = \xi \cdot \text{inf}_{\text{node}}(cu_k) + (1 - \xi) \cdot \text{inf}_{\text{clust}}(cu_k) \quad (10)$$

$$\text{inf}_{\text{node}}(cu_k) = \sum_{i=1}^I C_{i,k}^{(1)} \times I(u_i)$$

$$\text{inf}_{\text{clust}}(cu_k) = \sum_{q \neq k} D_{k,q} + \psi \cdot \sum_{l=1}^L B_{k,l}$$

where the node-level informativeness $\text{inf}_{node}(cu_k)$ is directly calculated by summing up the importance of the shots that belong to the cluster; the cluster-level informativeness $\text{inf}_{clust}(cu_k)$ is defined as a combination of its intra-type relations with other shot clusters and its inter-type relations with keyword clusters with a weighting factor ψ .

Finally, we select representative keyframes and keywords of hidden topics. Each cluster indicates a hidden sub-theme or sub-event. Therefore, we propose a two-step selection procedure to maximize the coverage of the topics while at the same time removing redundancy as more as possible.

- Select a number of shot clusters with highest importance score, and only choose the keyframe of the most important shot in each cluster as its representation.
- Select a number of keyword clusters with highest importance score, and only choose the most important keyword in each cluster as its representation.

2.4. Circular Storyboard Presentation

After mining topic-representative keyframes and keywords, we use a circle space to present the topic-representative keyframes and keywords as circular storyboard. The representative keyframes of visual topics are averagely placed at regular intervals on the boundary of the circle. The representative keywords of textual topics are displayed inside the circle at a position which is determined to their relevance with the shot cluster.

$$\text{point}(cw_l) = \sum_{k=1}^K \text{point}(cu_k) \times B_{k,l} \quad (11)$$

In order to guild user's attention to important topics, keyframes and keywords are resized according to their importance score, so that higher importance shot clusters are represented with larger keyframes, higher importance keyword cluster are represented with larger font size.

$$\text{size}(cw_l) = I(cw_l) / I(cw_l)_{\max} \quad (12)$$

$$\text{size}(cu_k) = 0.5 + 0.5 \times I(cu_k) / I(cu_k)_{\max}$$

Table 1. Test video set

No.	Category	Video num.	Shot num.	Total length
I-1	Sci-Tech	3	27	2:33
I-2	Sci-Tech	9	105	7:05
II-1	Business	10	117	9:02
II-2	Business	15	226	15:22

3. Experiments

3.1. Data set

We have collected a corpus of CCTV Broadcast News during January 1, 2008 and May 31, 2008 from CCTV News Broadcasting Program website (<http://news.cctv.com/program/xwlb>). Speech transcript for each video was also collected using a large vocabulary continuous speech recognition engine [13]. The corpus contains a total of 2080 video clips with wide diversity in topics. Following the news classification scheme of Baidu News site (<http://news.baidu.com/>), we use transcript text to classify videos into 13 categories: world, China, sports, entertainment, society, business, internet, science & technology (Sci-Tech), house, auto, culture, education, health and game. And then we employ an affinity propagation clustering algorithm [14] to group videos in each category into clusters. Each cluster contains a set of videos having a main topic. Four clusters from Sci-Tech and Business categories are chosen as our test set. Table 2 shows the cluster information including: the category, the number of videos, the total number of shots and the total length of each cluster.

3.2. Results and Evaluations

In our proposed approach, we formulated multi-video summarization as a complex graph clustering and mining problem and represented the selected representative keyframes and keywords as an image-plus-text circular storyboard. To compare the performance with our proposed method, we implement two storyboard approaches as the baselines. As for one of them, baseline-I, we simplify the complex graph clustering and mining algorithm into an extremely special case by setting all elements of affinity matrix A described in Sub-Section 3.2 as zero to ignore the relations of shots and keywords. The representative keywords with highest importance score are directly selected; the representative keyframes with most visual informativeness are selected from clusters grouped from a simplified homogeneous graph clustering. And the finally selected keyframes and keywords are visualized as a list storyboard by placing keyframes in upper half and keywords in lower half. The goal of baseline-I is to evaluate the effectiveness of our proposed complex graph clustering and mining algorithm. As for another storyboard system, baseline-II, we just replace circular storyboard presentation in our proposed approach as a list storyboard presentation while keeping the others not changed. The goal of baseline-II is to evaluate the presentation quality of circular storyboard.

In order to quantitatively evaluate the effectiveness of our proposed approach, we evaluate informativeness between the two baselines and our

proposed one. The criterion informativeness measures whether the summary can bring users as much as the original video set do or not. We invited 12 graduate students, including 8 males and 4 females, to give subjective scores ranging from 1 (worst) to 100 (best) to the criterion. After browsing the three types of video summaries, they can watch the original video and write down their decisions.

Table 2 lists the results. Compared to baseline-I with an average score of 69.5, baseline-II achieves an average score of 74.1 with a relative improvement of 6.6%. These results demonstrate the effectiveness of proposed complex-graph clustering and mining algorithm. In addition, compared to baseline-II with an average score of 74.1, our proposed approach obtain an average score of 77.6 with a relative improvement of 4.7% which indicates that the presentation of summary as circular storyboard can offer more information to users.

Table 2. Evaluation results

Video set	Baseline-I	Baseline- II	Proposed
I-1	82.3	84.8	86.8
I-2	66.0	73.1	77.4
II-1	71.6	78.7	83.3
II-2	58.1	59.8	62.7
Avg.	69.5	74.1	77.6

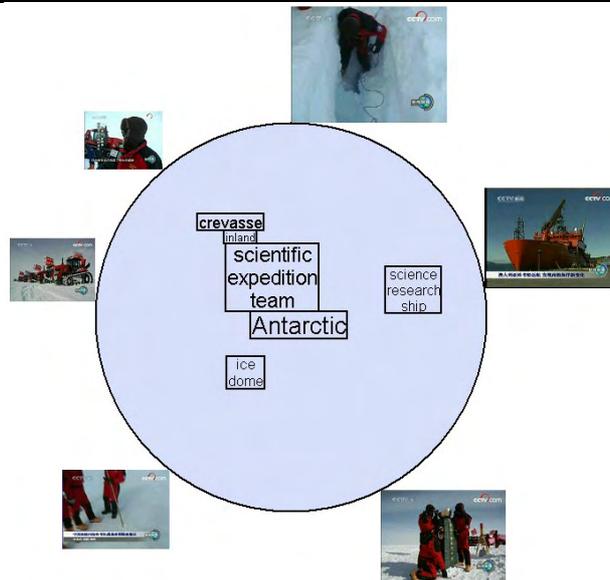


Fig. 2. Circular storyboard based on complex graph clustering and mining

Figure 2 illustrate the example of the summarized results using our approach, in video set 1-1 which has a main topic “science expedition in

Antarctic” (Note that the original Chinese keywords have been mapped into English keywords for illustration and explanation).

4. Conclusions

Multi-video summarization faces technical challenges due to the wider diversity of topics in multi-video than single-video as well as the multi-modality nature of multi-video over multi-document. We propose techniques to analyze both visual and textual content across a set of videos and to create a circular storyboard composed of topic-representative keyframes and keywords. We formulate the generation of circular storyboard as a complex graph clustering and mining problem, in which the divided shots from visual data and the extracted keywords from speech transcripts are first structured into a complex graph and grouped into clusters; the representative keyframes and keywords of hidden topics are then mined from clustered complex graph to maximize the coverage in topics while at the same time removing redundancy as more as possible. Experiments carried out on four video clusters of CCTV Broadcast News videos show the effectiveness of our proposed approach.

5. Acknowledgement

This work is supported by National Natural Science Foundation of China (No. 60833006), National Key Technology R&D Program (2007BAH11B01), Science and Technology Project of Zhejiang Province (2008C13G1410001, 2008C13075, 2008C13081, 2008C13G2010039)

6. References

1. Christel M. G. and Hauptmann A. G. and Lin W. H. and Chen M. Y. and Yang J. and Maher B. and Baron R. V.: Exploring the utility of fast-forward surrogates for BBC rushes. Proc. TRECVID Summarization Workshop, Vancouver, British Columbia, Canada. (2008)
2. Christel M. G.: Evaluation and user studies with respect to video summarization and browsing, IS&T/SPIE Symposium on Electronics Imaging, San Jose, CA. (2006).
3. Yeung, M.M. and Yeo, B.L.: Video visualization for compact presentation and fast browsing of pictorial content. IEEE Trans. Circuits and System for Video Technology, vol. 7, no. 5, 771-785. (1997)9.
4. Chen, B. W. and Wang, J. C. and Wang, J. F.: A novel video summarization based on mining the story-structure and semantic relations among concept entities. IEEE Tran. Multimedia, vol. 11, no. 2. (2009)
5. Ding, W.: Multimodal surrogates for video browsing. Proc. ACM Digital Lib, Berkeley, CA, 85-93. (1999)

Jian Shao, Dongming Jiang, Mengru Wang, Hong Chen, and Lu Yao

6. Wan, X.J. and Yang, J.W.: Improved affinity graph based multi-document summarization. Proc. Human Language Technology Conf. of the NAACL, 181-184. (2006)
7. Wang F. and Merialdo B.: Multi-document video summarization, Proc. ICME'09. (2009)
8. Liu Y. N., Wu F., Zhuang Y. T., Xiao J.: Active post-refined multi-modality video semantic concept detection with tensor representation, ACM Multimedia 2008, 91-100. (2008)
9. Long, B. and Zhang, Z. F. and Yu P. S.: Clustering on Complex Graph, Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence. (2008)
10. Ye, Z.Y. and Wu, F., A robust fusion algorithm for shot boundary detection. Journal of Computer Aided Design and Computer Graphics (In Chinese with English Abstract), vol.15, No.11, 950-955. (2003)23.
11. Odobez, J.M. and Gatica-Perez, D.: Video shot clustering using spectral methods. Proc. 3rd Int. Workshop on Content-Based Multimedia Indexing, Rennes, France, 94-102. (2003)
12. Deodhar M. and Ghosh J. and Gupta G. and Cho H. and Dhillon I.: A scalable framework for discovering coherent co-clusters in noisy data, ICML'09. (2009)17.
13. Young S. and Everman G and et al.: The HTK book (for HTK version 3.4), <http://htk.eng.cam.ac.uk>. (2007)
14. Xia D.Y. and Wu F. and Zhang X.Q and Zhuang Y.T.: Local and global approaches of affinity propagation clustering for large scale data. Journal of Zhejiang University SCIENCE A, vol. 9 no. 10 pp. 1373~1381. (2008)

Jian Shao has received his B.S. degree in electrical science and engineering from Nanjing University in 2003, Ph.D. degree from Institute of Acoustics, Chinese Academy of Sciences in 2008. He is currently a post-doctor of College of Computer Science, Zhejiang University. His research is focused on multimedia analysis and retrieval.

Dongming Jiang has received his B.S. degree in computer science and technology from Zhejiang University (ZJU), China, in 2007. He is currently pursuing the M.S. degree in the computer application technology at ZJU. His research interests include video analysis and information retrieval.

Mengru Wang has received his B.S. degree in electrical science and engineering from Zhejiang University in 1994. Now he is a senior engineer and team leader of Zhejiang radio & TV Group. His research is focused on video processing and transmission

Hong Chen has received his M.S. degree in electrical science and engineering from Zhejiang University in 2005. Now he is an engineer of Zhejiang radio & TV Group. His research is focused on video processing and transmission.

Yao Lu has received his B.S. degree in software engineering from Zhejiang University (ZJU), China, in 2008. He is currently pursuing the M.S. degree in computer science at ZJU. His research interests include video processing and personalized recommendation.

Received: May 10, 2009; Accepted: September 21, 2009.

A Retrieval Method for Human Mocap Data Based on Biomimetic Pattern Recognition

Xiaopeng Wei¹, Boxiang Xiao¹, and Qiang Zhang¹

¹Key Laboratory of Advanced Design and Intelligent Computing(Dalian University),
Ministry of Education, Dalian, 116622, China
zhangq26@126.com

Abstract. A retrieval method for human Mocap (Motion Capture) data based on biomimetic pattern recognition is presented in this paper. BVH rotation channels are extracted as features of motion for both the retrieval instance and the motion data. Several hyper sausage neurons are constructed according to the retrieval instance, and the trained domain covered by these hyper sausage neurons can be considered as the distribution range of a same kind of motions. By use of CMU free motion database, the retrieval algorithm has been implemented and examined, and the experimental results are illustrated. At the same time, the main contributions and limitations are discussed.

Keywords: MoCap (Motion Capture) Data, Retrieval, Biomimetic Pattern Recognition, HSN (Hyper Sausage Neurons).

1. Introduction

In past decades, with the development and popularity of motion capture device, human motion capture including motion data processing and retrieval has attracted increasing attention from many researchers [1-15]. There is a rapidly growing data of human motion, and the reuse of human motion capture data is becoming a practical task. However, because of the lack of general and efficient motion retrieval systems, the reuse of the motion data is limited by many problems and the investigation of motion data retrieval approaches is still a hot issue.

In this paper, we present a retrieval method for human Mocap data based on biomimetic pattern recognition [16], and the method constructs a set of the HSN (Hyper Sausage Neurons) for each kind of motion. Biomimetic pattern recognition also called topological pattern recognition is a novel pattern recognition principles proposed by Wang and colleagues [17] which is based on "matter cognition" instead of "matter classification" in traditional statistical pattern recognition. This recognition model is better closer to the function of human being rather than traditional statistical pattern recognition using "optimal separating" as its main principle. Biomimetic pattern recognition covers the distribution of a feature vectors exactly in high-dimensional space,

and has been widely used in face recognition [18], speaker identify [19] and other feature recognition fields.

The aim of this work is to implement an effective retrieval approach for human motion capture data. The flowchart of method in this paper is shown in Fig1. The remainder of the paper is organized as follows: We introduce the feature extraction of motion data in Section 3 and describe our retrieval algorithm based on biomimetic pattern recognition in Section 4. Moreover, in Section 5, the part of experimental results and comparison to other methods are presented and illustrated. Finally, we conclude and explain the main advantages and limitations of this method and discuss the future works in Section 6.

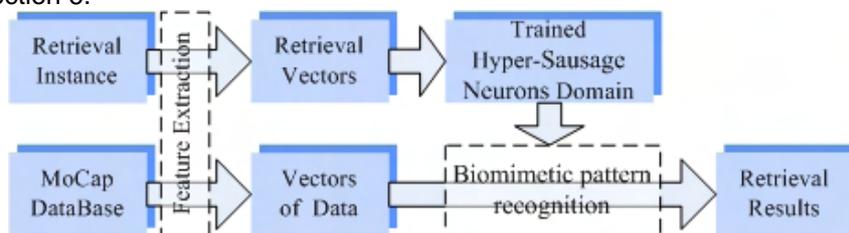


Fig. 1. The flowchart of method in this work.

2. Related Works

Many works in human motion data retrieval have been developed in past decades. In previous studies, one of developed techniques is motion templates based methods for human motion retrieval and classification. Muller and his colleagues [3] proposed a method for automatic classification and retrieval of motion capture data facilitating the identification of logically related motions scattered in some database. Roder introduced templates methods systematically in his doctoral dissertation [4]. Another technique is content-based or index-based methods. For instance, Chiu et.al [2] put forward to a framework for constructing a content-based human motion retrieval system including two major components: indexing and matching. Muller et al [5] presented automated methods for efficient indexing and content-based retrieval of motion capture data. Yamasaki and his colleagues [6] described a content-based cross search scheme for two kinds of three-dimensional (3D) human motion data: time-varying mesh (TVM) and motion capture data. Chao and colleagues [7] presented a simple and effective approach for motion retrieval and synthesis based on posture feature indexing, and posture features of each frame data was extracted by an index function. A 3D motion retrieval method with motion index tree was presented by Liu and colleagues [8]. Feature analysis is also a common used approach by many researchers. A set of relational motion features was been defined in Demuth's work [1]. Xiang and Zhu [9] extracted 3D temporal-spatial features of motion data and

automatically constructed data driven decision trees. Lin [10] also defined a kind of motion features in his work. Many researchers used dynamic time warping (DTW) for different extent motion sequences [1,2]. Furthermore, there are several other methods such as probabilistic principal component analysis (PPCA) proposed by Wang [11], movement notation language presented by Yu et.al [12], Ensemble HMM Learning based approach developed by Xiang and Zhu [13], energy morphing based method proposed by Tam et.al [14] and semantic matching based method [15].

3. Feature Extraction of Motion Data

BVH rotation channels are used as features of motion for both the retrieval instance and the motion data in this work. The BVH file format for motion capture is originally developed by Biovision [20], a motion capture services company, as a way to provide motion capture data to their customers. A BVH file has two parts, a header section which describes the hierarchy and initial pose of the skeleton; and a data section which contains the motion data [21]. In BVH format, all joints' position can be calculated by their all parents' local rotation channels and root's translation. The BVH format doesn't account for scales so it isn't necessary to worry about including a scale factor calculation.

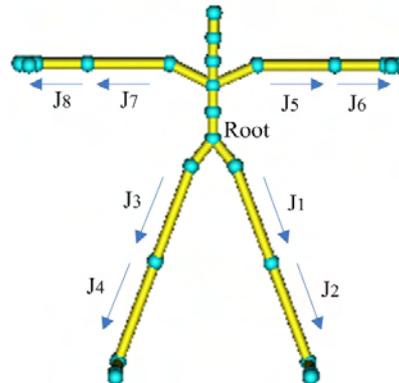


Fig. 2. 8 selected joints including Hips, Knees, Shoulders and Elbows

The BVH conversion of CMU motion data [22] we used in experiments is a standard database and all motions are represented by uniform BVH tree and scales. Here we select 8 joints including Hips, Knees, Shoulders and Elbows as the main feature of motion and extract their correlative values of local rotation channels in Motion sections of BVH files, as shown in Fig.2. Then, we define a 24 dimensional feature vector

$M^{24} = (Zrot_{J1}, Yrot_{J1}, Xrot_{J1}, Zrot_{J2}, \dots, Xrot_{J8})$ by the selected 8 joints' channel values, and all motions and retrieval instances are simplified by feature vectors.

4. Biomimetic Pattern Recognition-Based Retrieval Method

4.1. The Retrieval Instance

The retrieval instance in this work is a set of short query motion clips. Several frames of key pose in query clips are selected by manual operation, and all key pose frames' feature vectors mentioned in Section 3 are subsequently extracted. And then, a Remark Sequence of retrieval instance is constructed at the same time. Fig.3 shows an example of one motion clip of walk, in that the 4 heavy black frames are key poses of a walk clip with Remark Sequence [1,2,3,4], and others are general data.

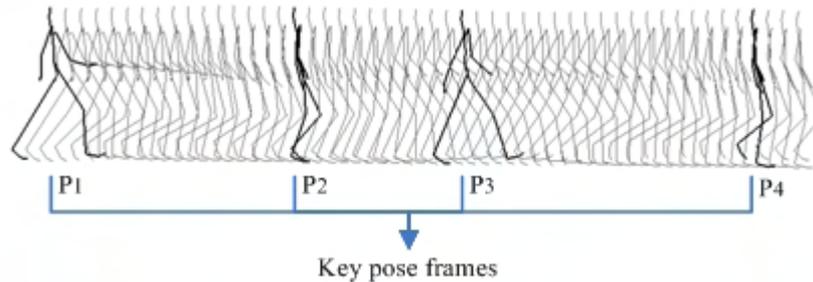


Fig. 3. Key poses of walk (heavy black)

4.2. HSN Covering in High-Dimensional Space

Biomimetic pattern recognition is a novel pattern recognition principles [17][18] which is based on “matter cognition” instead of “matter classification” in traditional statistical pattern recognition. Fig.4 shows the main principle of HSN chains based BPR and the comparison with traditional BP and RBF networks. The triangles represent samples to be recognized, and the circles and crosses represent samples to be distinguished from triangles. Polygonal line denotes the classification boundaries of traditional BP networks. Big circle denotes Radial basis function (RBF) networks.

A hyper sausage neuron can be expressed by topological product of a hyper-sphere and a line segment in high-dimensional space, as shown in Fig.5. For biomimetic pattern recognition, the task of geometrical learning is to cover a given sample set by a chain of hyper sausage units with a minimum sum of volumes via determining the end points of each line segment and the radius of each hyper-sphere.

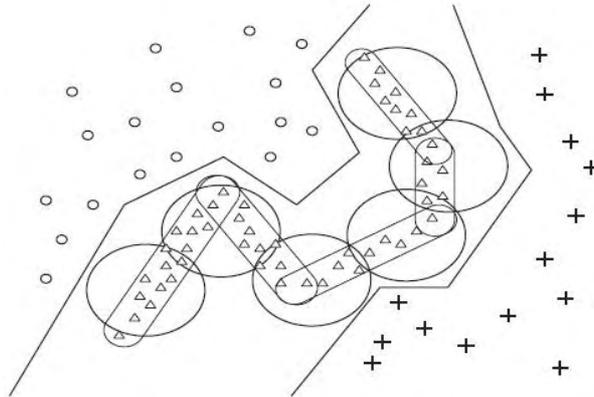


Fig. 4. HSN chains based BPR and comparison with BP and RBF networks.

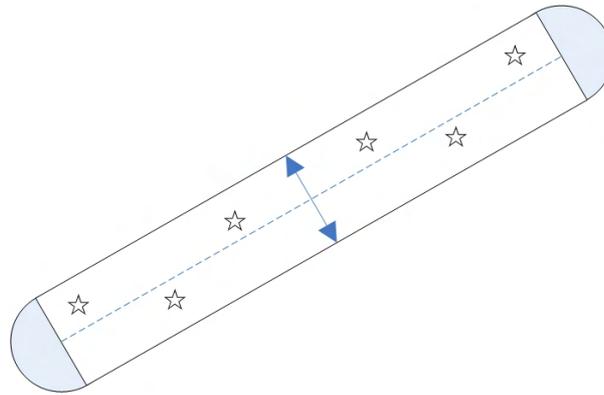


Fig. 5. A hyper sausage neuron schematic diagram in some 2D section.

And a hyper sausage neuron can be approximately expressed by Eq. 4.1-4.3:

$$f_{HSN}(X) = \text{sgn}\left(2 \frac{d^2(X, \overline{X_1, X_2})}{r^2} - 0.5\right) \quad 4.1$$

which contains a radius parameter r and the distance d between X and the line segment $\overline{X_1, X_2}$ as follows:

$$d^2(X, \overline{X_1, X_2}) = \begin{cases} \|X - X_1\|^2, & q(X, X_1, X_2) < 0, \\ \|X - X_2\|^2, & q(X, X_1, X_2) > \|X_1 - X_2\|, \\ \left\| \|X - X_1\|^2 - q^2(X, X_1, X_2) \right\|, & \text{otherwise} \end{cases} \quad 4.2$$

$$q(X, X_1, X_2) = (X - X_1) \frac{(X_1 - X_2)}{\|X_1 - X_2\|} \quad 4.3$$

In motion retrieval, we consider a feature vector of a frame of motion clip in motion database as X , and calculate the distances between X and all

feature vectors of a kind of key pose with same mark. And then, we note the two nearest objects in retrieval instance as X_1 and X_2 respectively. Finally, we apply the construction algorithm of hyper sausage neuron with the parameters X , X_1 , X_2 and a threshold r by Eq4.1-4.3. If the sign of calculating result $f_{HSN}(X)$ is plus, the feature vector X can be considered as a similar object to the kind of key pose, and be remarked by same mark of the key pose.

4.3. The Evaluation of Similarity

For motion retrieval, hyper sausage neurons covering in high-dimensional space is used to estimate the distribution of every motion data in the space. If a frame's feature vector in a motion clip have a same mark to pose P_k , it can be considered as a similar object to P_k and be noted as k in Cover Sequence, or else it is noted as 0 . Cover Sequence noted as C is an integer array to record the similar status of a motion sequence, for example in Eq4.4, and it can be simplified. In sequence, zero elements are removed, and continuous same elements are denoted by one element.

$$C = [0, 0, 0, 1, 1, 0, 2, 2, 2, 2, 0, 0, 3, 4, 4]$$

$$C_{simple} = [1, 2, 3, 4]$$

4.4

The Cover Sequence of a motion sequence is the unique evaluation for retrieval. If a motion's the Cover Sequence continuously cover the Remark Sequence in order, it is considered as the similar object. Furthermore, part sequence of a sequence is defined to evaluate main similarity, and it is actually the sub sequence of a sequence with one element absent. For example, sequences $[1, 2, 3]$, $[2, 3, 4]$, $[1, 2, 4]$, $[1, 3, 4]$ are all part sequences of sequence $[1, 2, 3, 4]$.

For instance, if in a Cover Sequence C_{walk} of a walk sequence, a sub sequence equal to the Remark Sequence of retrieval instance such as $[1, 2, 3, 4]$ in Fig.3 exists, the motion can be considered as the similar object to walk and it can be retrieved. If the Cover Sequence of a motion contains part but whole Remark Sequence of retrieval instance, we can consider it as the main similar object which can also be recalled.

5. Experimental Results

To examine our algorithm, we developed a motion retrieval system by C++ program language and OpenGL (Open Graphic Library), and all experimental

results were obtained on a 3.0 GHz Pentium 4 with 2 GB of main memory. We evaluated the system on a subset of BVH conversion of the CMU motion database [22], which contains 184 motions about 105 thousands frames sampled at 120 Hz of motion capture data including walk, run, jump, cartwheel, swing and their blends.

The retrieval accuracy of the proposed framework is evaluated by the precision and recall which are commonly adopted by many researchers.

$$\text{precision} = \frac{\#\{\text{relevant} \cap \text{retrieved}\}}{\text{retrieved}}$$

$$\text{recall} = \frac{\#\{\text{relevant} \cap \text{retrieved}\}}{\text{relevant}}$$

where #retrieved is the number of retrieved clips and #relevant is the number of relevant clips. Table 1 shows the part of retrieval results for walk and jump.

Table 1. part of retrieval results for walk and jump

	threshold	precision	recall
walk	30.0	1.000	0.862
	40.0	0.930	0.914
	45.0	0.688	0.914
	50.0	0.578	0.948
	55.0	0.586	1.000
jump	50.0	1.000	0.562
	55.0	0.983	0.795
	60.0	0.848	0.918
	70.0	0.634	0.973
	75.0	0.598	1.000

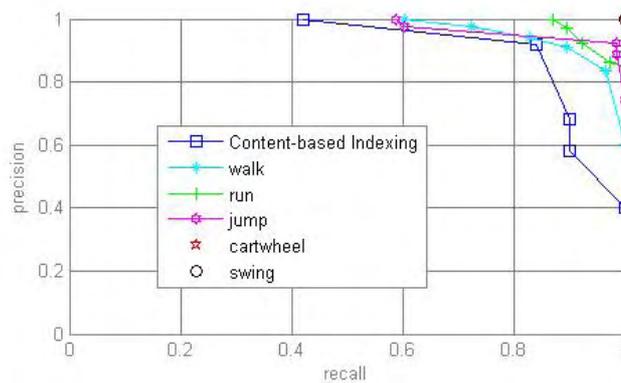


Fig. 6. part results of retrieval accuracy of proposed approach and the comparison to Content-based Indexing method.

The part results are also shown in Fig. 6, and Fig.7 shows part visual retrieval results. In experiments, some motions such as cartwheel and swing achieve results with accuracies at 1.0, because the numbers of samples in

database are few respectively at 9 and 10. To evaluate the performance of ours method, we compare our accuracy with the referenced Content-based Indexing method [2], shown in Fig 6.

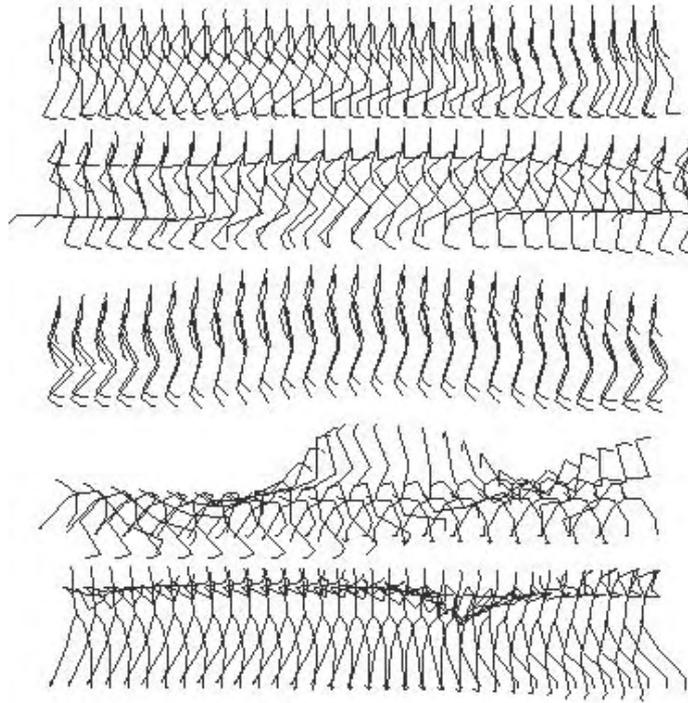


Fig. 7. part visual retrieval results including walk, run, jump, cartwheel, swing from upper to lower.

- On the other hand, time cost of approach is evaluated in ours work. The time cost includes two main parts: feature vectors extraction and computation of HSN covering. A main time-consuming operation is feature vectors extraction for the whole database (about 2 minutes) and the operation needs to be executed only one time in all experiments because the structures of feature vectors are same and uniform for different retrieval instances. Moreover, the time cost of covering computation traversing the whole database about for a retrieval instance is about 25 seconds.

6. Conclusion

In this paper, we presented a retrieval method for human Mocap (Motion Capture) data based on biomimetic pattern recognition is presented in this paper. The retrieval instance in this work is a set of short query motion clips. BVH rotation channels are extracted as features of motion for both the

retrieval instance and the motion data. Several hyper sausage neurons are constructed according to the retrieval instance, and the trained domain covered by these hyper sausage neurons can be considered as the distribution range of a same kind of motions. By computation of hyper sausage neurons covering based on biomimetic pattern recognition, relevant motions to the retrieval instance are recalled, and experimental results are illustrated.

The advantage of this work is twofold. Firstly, the proposed algorithm is based on the distribution of similar motions in high-dimensional space, and avoids the influences by differences in extent and velocity of motion sequences. Furthermore, the use of BVH rotation channels of motion data eliminates the differences in geometric parameters including coordinates, position, direction and scales of motions, and the uniform structure of feature vectors ensure the feasibility and efficiency of retrieval in large database.

One limitation of our approach is that the covering domains of hyper sausage neurons of some similar motions such as walk and run may be intersectant in few scenes which can reduce the retrieval accuracy. Another shortage is that some complex freeform motion can not be expressed by distribution in multidimensional space well and it is a difficult task for future work. Moreover, we are also focus on the dynamic distribution of motions in high-dimensional space.

7. Acknowledgment

This work is supported by National Natural Science Foundation of China (Grant No. 60875046), by Program for New Century Excellent Talents in University (Grant No. NCET-06-0298), by the Key Project of Chinese Ministry of Education.(No.209029), by China Postdoctoral Science Foundation ((Grant No. 2009045 1266), by the Program for Liaoning Innovative Research Team in University (Grant Nos.2008T004 and 2009T005), by the Program for Liaoning Science and Technology Research in University (Grant Nos.2009S008,2009S009) and by the Program for Liaoning Excellent Talents in University (Grant No.2006R06).

8. References

1. Demuth, B. and Roder, T. and Muller, M. and Eberhardt, B: An Information Retrieval System for Motion Capture Data. LNCS 3936. 373–384. (2006)
2. Chiu, C. and Chao, S. and Wu, M. and Yang, S. and Lin, H.: Content-based retrieval for human motion data. Journal of Visual Communication and Image Representation. 3, 15, 446–466. (2004) Muller, M. and Roder, T.: Motion Templates for Automatic Classification and Retrieval of Motion Capture Data. Eurographics/ ACM SIGGRAPH Symposium on Computer Animation. 137-146. (2006)

Xiaopeng Wei, Boxiang Xiao, and Qiang Zhang

3. Roder, T.: Similarity, Retrieval, and Classification of Motion Capture Data. University of Bonn. 93-110. (2006)
4. Muller, M. and Roder, T. and Clausen, M.: Efficient Content-Based Retrieval of Motion Capture Data. ACM Transactions on Graphics (TOG). 24, 3, 677 - 685. (2005)
5. Yamasaki, T. and Aizawa, K.: Content-Based Cross Search for Human Motion Data Using Time-Varying Mesh and Motion Capture Data. Proceeding of IEEE International Conference on Multimedia and Expo (ICME2007). 2007-2009. (2007)
6. Chao, S. and Chiu, C. and Chao, J. and Ruan, Y. and Yang, S.: Motion Retrieval and Synthesis Based on Posture Feature Indexing. Proceedings of the Fifth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'03). 266 - 271. (2003)
7. Liu, F. and Zhuang, Y. and Wu, F. and Pan, Y.: 3D motion retrieval with motion index tree. Computer Vision and Image Understanding. 92, 2, 265-284. (2003)
8. Xiang, J. and Zhu, H.: Motion Retrieval Based on Temporal-Spatial Features by Decision Tree. LNCS 4561. 224-233. (2007)
9. Lin, Y.: Efficient Motion Search in Large Motion Capture Databases. LNCS 4291. 151-160. (2006)
10. Wang, X. and Yu, Z. and Wong, H.: 3D Motion Sequence Retrieval Based on Data Distribution. ICME 2008. 1229-1232. (2008)
11. Yu, T. and Shen, X. and Li, Q. and Geng, W.: Motion retrieval based on movement notation language. Computer Animation and Virtual Worlds. 16, 273-282. (2005)
12. Xiang, J. and Zhu, H.: Ensemble HMM Learning for Motion Retrieval with Non-linear PCA Dimensionality Reduction. IJHMSP 2007. 604 - 607. (2007)
13. Tam, G. and Zheng, Q. and Rynson, M. and Lau, W.H.: Motion Retrieval Based on Energy Morphing. Ninth IEEE International Symposium on Multimedia 2007. 210-217. (2007)
14. Wei, X. and Xiao, B. and Zhang, Q. and Zhou, D.: Effective approach based on semantic matching for retrieval of MoCap data. Journal of Computational Information Systems. 3, 4, 859-866. (2008)
15. Wang, S. and Lai, J.: Geometrical learning, descriptive geometry, and biomimetic pattern recognition. Neurocomputing. 67, 9-28. (2005)
16. Wang, S.: Bionic (Topological) Pattern Recognition-A New Model of Pattern Recognition Theory and Its Applications. Chinese Journal of Electronics. 30, 10, 1417-1420. (2002)
17. Wang, Z. and Mo, H. and Lu, H. and Wang, S.: A method of biomimetic pattern recognition for face recognition. Proceedings of the International Joint Conference on Neural Networks 2003. 2216-2221. (2003)
18. Qin, H. and Wang, S. and Sun, H.: Biomimetic Pattern Recognition for Speaker-Independent Speech Recognition. International Conference on Neural Networks and Brain 2005. 1290-1294. (2005)
19. [http://www. Biovision.com](http://www.Biovision.com).
20. <http://www.cs.wisc.edu/graphics/Courses/cs-838-1999/Jeff/BVH.html>
21. <http://mocap.cs.cmu.edu/>

Xiaopeng Wei is a professor at Dalian University, Dalian, China. His research area include computer animation, intelligent CAD. So far, he has (co-) authored about 160 papers published.

A Retrieval Method for Human Mocap Data Based on Biomimetic Pattern Recognition

Boxiang Xiao is a lecturer at Dalian University, Dalian, China. He has received his Ph.D degree in 2009 at Dalian University of Technology.

Qiang Zhang is a professor at Dalian University, Dalian, China. His research interests are computer animation, intelligent computing. Now he has served as editorial boards of seven international journals and has edited special issues in journals such as Neurocomputing and International Journal of Computer Applications in Technology.

Received: May 15, 2009; Accepted: October 26, 2009.

Generative 3D Images in a visual evolutionary computing system

Hong Liu

School of Information Science and Engineering, Shandong Normal University, Jinan
lhscn@jn-public.sd.cninfo.net

Abstract. This paper presents a novel computer-aided design system which uses a computational approach to producing 3D images for stimulating creativity of designers. It introduces the genetic algorithm first. Then a binary tree based genetic algorithm is presented. This approach is illustrated by a 3D image generative example, which uses complex function expressions as chromosomes to form a binary tree, and all genetic operations are performed on the binary tree. Corresponding complex functions are processed by MATLAB software to form 3D images of artistic flowers. This generative design is integrated with a visualization interface, which allows designers to interact and select from instances for design evolution. It shows the system is able to enhance the possibility of discovering various potential design solutions.

KeyWords: Evolutionary computation, computer-aided design, generative design, complex function

1. Introduction

To date, more and more computers have been used in design offices. The CAD (computer-aided drafting or computer-aided design) tools frequently used in these design offices provide two types of support. These tools are for two-dimensional (2D) drafting and three-dimensional (3D) modeling tasks. The drafting tools replace traditional tools such as pencils and rulers to produce detailed design drawings, while modeling tools take over the functions to allow the visualization of new designs. These tools assist designers in the final production and presentation of the design products.

However, few of those CAD tools have been used to assist designers in the early phases, such as the conceptual design process. During the early phases, designers explore many design alternatives. Current CAD tools do not provide sufficient design support to innovative design.

In cognitive psychology, design activities are described as specific problem-solving situations, since design problems are both ill defined and open-ended. Design activities, especially in 'non-routine activities', designers involve a special thinking process. This process includes not only thinking with logic, but also thinking with mental imagery and sudden inspiration.

Designers have called new ideas in their mind as idea sketches. In contrast to presentation sketches, idea sketches are made in the early phases of design. They function as a tool to interact with imagery and are predominantly for private use. Because of their early appearance in the design process, idea sketching will have an important role in creative processes. This is the reason why many computer tools aim at supporting and improving idea sketching.

Creative ideas occur in a particular medium. Most of the researchers in the field of creativity agree that designers who are engaged in creative design tasks use external resources extensively. Such external resources include a variety of physical and logical information. For instance, reading books, browsing photographic images, talking to other people, listening to music, looking at the sea or taking a walk in the mountains. Sketches and other forms of external representations produced in the course of design are also a type of external resources that designers depend on. When designers discover a new or previously hidden association between a certain piece of information and what they want to design, the moment of creative brainwave emerges. Designers then apply the association to their design and produce an innovative design.

Visual images are particularly for activating creativity. In product design, visual expression, especially in the form of sketching, is a key activity in the process of originating new ideas. This approach suffers from the fact that most creative processes extensively make use of visual thinking, or, in other words, there is a strong contribution of visual imagery. These processes are not accessible to direct verbalization.

The design research community has spent much of its effort in recent years developing design systems for supporting innovative design. Generative design systems - systems for specifying, generating and exploring spaces of designs and design alternatives - have been proposed and studied as a topic of design research for many years.

Generative design is an excellent snapshot of the innovative process from conceptual framework through to specific production techniques and methods. It is ideal for aspiring designers and artists working in the field of computational media, especially those who are interested in the potential of generative, algorithmic, combinational, emergent and visual methods as well as the exploration of active images.

This paper presents a novel computer supported design system that uses the evolutionary approach to generate 3D images. The tree structure based genetic algorithms and complex functions are used in this system. This generative design is integrated with a visualization interface, which allows designers to interact and select from instances for design evolution. Programs are implemented by using Visual C++6.0 and mathematical software MATLAB.

The remainder of this paper is organized as follows. Section 2 is concerned with related work on generative design. Section 3 introduces genetic algorithms and genetic programming. In section 4, a binary tree based genetic algorithm is presented. Section 5 illustrates an artwork design example for showing how to use the tree structure based genetic algorithm and complex

functions to generate 3D images. Finally, these results are briefly analyzed, followed by a discussion of possible future improvements.

2. Related work

Generative systems are relevant to contemporary design practice in a variety of ways. Their integration into the design process allows the development of novel design solutions, difficult or impossible to achieve via other methods. Grammar-based techniques exploit the principle of database amplification, generating complex forms and patterns from simple specifications. Evolutionary systems may be used in combination with aesthetic selection to breed design solutions under the direction of a designer. Interface design and other sign systems may be defined in terms of adaptive procedures to create communication that adapts to its interpretation and use by an audience [1].

The key properties of generative systems can be summarized as [2]:

- The ability to generate complexity, many orders of magnitude greater than their specification. This is commonly referred to as database amplification, whereby interacting components of a given complexity generate aggregates of far greater behavioral and/or structural complexity. Such aggregates may in turn generate their own interactions forming new aggregates of even higher sophistication and complexity. This is referred to as a dynamic hierarchy, a poignant example being complex multi-cellular organisms, whose hierarchy can be summarized: atom; molecule; organelle; cell; organ; organism; ecosystem.

- The ability to generate novel structures, behaviors, outcomes or relationships. Novelty used in this sense means the quality of being new, original and different from anything else before it. There are of course, different degrees of novelty. RNA and DNA was novel in that they introduced a completely new mechanism for replication and encoding of protein synthesis. Artists and designers are always seeking novelty (the opposite of which is mimicry or copying, something depreciated in the art and design world). Artistic novelty may not have such a significant impact as, for example, DNA, but the key concept is that of the new — generative systems have the potential to give rise to genuinely new properties. This is why they are often referred to as emergent systems. These new properties typically fall outside the designer's expectations or conceptualizations for the design, resulting in functionality or outcomes that were not anticipated. This of course raises the issue of control, a problematic issue for generative design, particularly if the designer is accustomed to organizing outcomes in a predictable way.

Generative design describes a broad class of design where the design instances are created automatically from a high-level specification. Most often, the underlying mechanisms for generating the design instances in some way model biological processes: evolutionary genetics, cellular growth, etc. These artificial simulations of life processes provide a good conceptual basis for designing products.

Evolutionary systems are based on simulating the process of natural selection and reproduction on a computer. This technique has found wide application in design for computer animation and graphics, but also in architectural, industrial and engineering design [3]. The technique depends on the specification of a parameterized model that is general enough to allow a wide variety of possible outcomes of interest to the designer. In cases where a very specific goal is sought, the parameter space must also be sufficiently broad to encapsulate an answer to the problem that meets the specified design constraints.

Initially a population of potential designs is generated with a random set of parameters. This random population may be displayed visually to the designer. The designer's aesthetic sense then determines the 'fittest' designs of those displayed, and these are 'bred' with one another to produce a new population of designs that inherit the traits of their successful parents. This process is akin to selective breeding of apple trees for the taste and color of their fruit – both subjective qualities assessed by humans [4].

An alternative means of utilizing this evolutionary process exists where a fitness function may be explicitly coded by the designer. For example, perhaps a designer seeks the lightest, cheapest set of automobile wheels. This fitness function is coded into the evolutionary system and the computer evolves populations of wheels towards a successful structure independently of further human input. Where creative designs are sought with some aesthetic value, the former technique of interactive evolution is more practical. This requires constant human interaction, a necessary bottleneck as long as it remains difficult to encode subjective qualities like 'beauty', 'ugliness' etc. in such a way that a computer can operate with them.

Some of the evolutionary design work was performed by Professor John Frazer, who spent many years developing evolutionary architecture systems with his students. He showed how evolution could generate many surprising and inspirational architectural forms, and how novel and useful structures could be evolved [5]. In Australia, the work of Professor John Gero and his colleagues also investigated the use of evolution to generate new architectural forms. This work concentrates on the use of evolution of new floor plans for buildings, showing over many years of research how explorative evolution can create novel floor plans that satisfy many fuzzy constraints and objectives [6]. They even show how evolution can learn to create buildings in the style of well-known architects.

In Argenia, a system for architectural design by Soddu, the three-dimensional models produced can be directly utilized by industrial manufacturing equipment like numerically controlled machines and robots, which already represent the present technologies of industrial production. This generative and automatic reprogramming device of robots makes it possible to produce unique objects with the same equipment and with costs comparable to those of objects that are identical; like a printer that can produce pages that are all the same or all different, at precisely the same cost [7].

Many artists and designers have turned to generative systems to form their design basis. Dextro (www.dextro.org) has developed a diverse range of

interactive drawing systems in which simple design elements such as points and lines replicate and self-organize to create illustrations and animations. Digital artists Meta use generative processes to create streams of abstract video (www.meta.am), expressive of the multi-layered, fluid mutable nature of electronic space. Jared Tarbell (www.levitated.net) has experimented with the intersection of generative systems, typography and graphic patterns in his experimental web design projects. Software such as Auto-Illustrator (www.auto-illustrator.com) and Autoshop (www.signwave.co.uk) combine generative systems with the image composition and editing functions of popular computer drawing programs to 'automatically' create new designs ready for use in projects. Groboto (www.groboto.com), a program that allows users to develop their own systems for growing 3D forms, makes these systems accessible to a wider audience by placing a GUI in front of the lines of code usually required to work with generative systems.

However, the development of generative design tools is still at its early stage. The research and development of design support tools using evolutionary computing technology are still in process and have huge potential for the development of new design technology.

3. Genetic algorithms and genetic programming

General GA (genetic algorithm) is a search algorithm based on the mechanisms of natural selection. They lie on one of the most important principles of Darwin: survival of the fittest. John Holland, in the 1970s, thought that he could incorporate in a computer algorithm such a technique, to solve different problems through evolution [8].

Given a specific problem to solve, the input to the GA is a set of potential solutions to that problem, encoded in some fashion, and a metric called a fitness function that allows each candidate to be quantitatively evaluated. These candidates may be solutions already known to work, with the aim of the GA being to improve them, but more often they are generated at random.

The GA then evaluates each candidate according to the fitness function. In a pool of randomly generated candidates, of course, most will not work at all, and these will be deleted. However, purely by chance, a few may hold promise - they may show activity, even if only weak and imperfect activity, toward solving the problem.

These promising candidates are kept and allowed to reproduce. Multiple copies are made of them, but the copies are not perfect; random changes are introduced during the copying process. These digital offspring then go on to the next generation, forming a new pool of candidate solutions, and are subjected to a second round of fitness evaluation. Those candidate solutions which were worsened, or made no better, by the changes to their code are again deleted; but again, purely by chance, the random variations introduced into the population may have improved some individuals, making them into better, more complete or more efficient solutions to the problem at hand. Again

these winning individuals are selected and copied over into the next generation with random changes, and the process repeats. The expectation is that the average fitness of the population will increase each round, and so by repeating this process for hundreds or thousands of rounds, very good solutions to the problem can be discovered.

Genetic algorithms have been gained recognition as solution techniques for many complex optimization problems [9], and have been applied to numerous conceptual and preliminary design studies [10, 11]. A genetic algorithm is not truly an optimization technique; rather it is a computational representation of processes involved in natural selection as observed in biological populations. They resemble natural evolution more closely than many other approaches because they are based on the mechanics of natural selection and natural genetics.

The GA consists of a number of elements. Once a problem has been identified, the elements can be broken down as follows:

- A representation of a potential solution;
- A population of chromosomes;
- A fitness function for evaluating the relative merit of a chromosome;
- A selection method;
- One or more operations for modifying the selected chromosomes (typically crossover and mutation);

These elements are then employed in an iterative process until a solution is found or a termination condition is met. An abstraction of a typical GA is given as follows:

Generate initial population, $G(0)$;
Evaluate $G(0)$; (apply fitness function)
 $t=1$;
Repeat
 Generate $G(t)$ using $G(t-1)$; (apply operators)
 Evaluate (decode($G(t)$));
 $t=t+1$;
Until solution is found or termination.

A search algorithm balances the need for exploration -- to avoid local optima, with exploitation -- to converge on the optima. Genetic algorithms dynamically balance exploration versus exploitation through the recombination and selection operators respectively. With the operators as defined earlier, the schema theorem proves that relatively short, low-order, above average schema are expected to get an exponentially increasing number of trials or copies in subsequent generations [12]. Mathematically

$$m(h, t + 1) \geq \frac{m(h, t)f(h)}{f_t} \left[1 - P_c \frac{\delta(h)}{l-1} - O(h)P_m \right]$$

Here $m(h,t)$ is the expected number of schemas h at generations t , $f(h)$ is the fitness of schema h and f_t is the average fitness at generation t . The genotype length is l , $\delta(h)$ is the defining length and $O(h)$ is the order of

schema h . P_c and P_m are the probabilities of crossover and mutation respectively.

The schema theorem leads to a hypothesis about the way genetic algorithms work.

Genetic algorithms traditionally use string-based representations in their chromosomes, but this is not always suitable for representing higher level knowledge. Koza introduced a hierarchical GA approach [13], where his chromosomes are tree-like expressions that can be recombined by swapping sub-trees. It starts with a randomly generated population of function-trees, the trees which perform best on the problem in question are selected to be the breeding stock. These parent trees are combined (by exchanging sub-trees) and mutated (by generating new sub-trees) to produce new trees for the next generation of the population which inherit some features from their parents. The next generation of trees is then evaluated against the problem. The best trees are selected to produce the next generation after that, and so on.

Koza labeled his hierarchical version of GA, the Genetic Programming Paradigm (called GP). The main difference between traditional GA and GP lies in the following graph theoretical manipulation.

1. Mutation changing a node label, or substituting sub-trees.
2. Crossover i.e. swapping sub-trees

As with genetic algorithm, the GP population is evolved to (hopefully) produce function-trees which can perform well on the problem in question. GP has been used successfully in generating computer programs for solving a number of problems in a wide range of areas [14].

Nowadays, genetic programming is applied mostly related to adaptive system and optimization, where representation of programs is used in conjunction with hybrid crossover to evolve a multiplication function. In addition, the design with genetic programming is not traditionally considered in canonical genetic programming.

4. The tree structured genetic algorithm

General genetic algorithms use binary strings to express the problem. It has solved many problems successfully. But it would be inappropriate to express flexible problem. For example, mathematical expressions may be of arbitrary size and take a variety of forms. Thus, it would not be logical to code them as fixed length binary strings. Otherwise, the domain of search would be restricted and the resulting algorithm would be restricted and only be applicable to a specific problem rather than a general case. Thus, tree structure, a method useful for representing mathematical expressions and other flexible problems, is presented in this paper.

The main contribution of this paper is to use tree-like expressions as chromosomes for representing complex mathematical functions and apply the tree structured genetic algorithm in generative design.

For a thorough discussion about trees and their properties, see [15,16]. Here, we only make the definitions involved in our algorithm and these definitions are consistent with the basic definitions and operations of the general tree.

Definition 1 A binary complex function expression tree is a finite set of nodes that either is empty or consists of a root and two disjoint binary trees called the left sub-tree and the right sub-tree. Each node of the tree is either a terminal node (operand) or a primitive functional node (operator). Operand can be either a variable or a constant. Operator set includes the standard operators (+, -, *, /, ^), basic mathematic functions (such as sqrt(), exp(), log()), triangle functions (such as sin(x), cos(x), tan(x), cot(x), sec(x), csc(x), asin(x), acos(x)), hyperbolic functions (such as sinh(), cosh(), tanh(), asinh(), acosh(), atanh()), complex functions (such as real(z), imag(z), abs(z), angle(z), conj(z)) and so on.

Here we use the expression of mathematical functions in MATLAB (mathematical software used in our system).

A binary complex function expression tree satisfies the definition of a general tree.

Genetic operations include crossover, mutation and selection. According to the above definition, the operations are described here. All of these operations take the tree as their operating object.

(1) Crossover

The primary reproductive operation is the crossover operation. The purpose of this is to create two new trees that contain 'genetic information' about the problem solution inherited from two 'successful' parents. A crossover node is randomly selected in each parent tree. The sub-tree below this node in the first parent tree is then swapped with the sub-tree below the crossover node in the other parent, thus creating two new offspring.

(2) Mutation

The mutation operation is used to enhance the diversity of trees in the new generation, thus opening up new areas of 'solution space'. It works by selecting a random node in a single parent and removing the sub-tree below it. A randomly generated sub-tree then replaces the removed sub-tree.

(3) Selection

For general design, we can get the requirement from designer and transfer it into goal function. Then, the fitness value can be obtained by calculating the similar degree between the goal and individual by a formula. However, for creative design, there are no standards to form a goal function. Therefore, it is difficult to calculate the fitness values by a formula. In our system, we use the method of interaction with the designer to obtain fitness values. The range of fitness values is from -1 to 1. After an evolutionary procedure, the fitness values appointed by designer are recorded in the knowledge base for reuse. Next time, when the same situation appears, the system will access them from the knowledge base [17].

This method gives the designer the authority to select their favored designs and thus guide the system to evolve the promising designs. Artificial selection

can be a useful means for dealing with ill-defined selection criteria, particularly user centered concerns.

Many explorative systems use human input to help guide evolution. Artists can completely take over the role of fitness function. Because evolution is guided by human selectors, the evolutionary algorithm does not have to be complex. Evolution is used more as a continuous novelty generator, not as an optimizer. The artist is likely to score designs highly inconsistently as he/she changes his/her mind about desirable features during evolution, so the continuous generation of new forms based on the fittest from the previous generation is essential. Consequently, an important element of the evolutionary algorithms used is non-convergence. If the populations of forms were ever to lose diversity and converge onto a single shape, the artist would be unable to explore any future forms.

For clarity, we will present the performing procedure of the tree structured genetic algorithms together with a flowers generative design example, in the next section.

5. A generative artwork example

An artwork design example is presented in this section for showing how to use tree structure based genetic algorithm and complex function expressions to generate 3D images.

The complex function expressions are used to produce 3D artistic images. Here, $z=x+iy$ (x is real part and y is virtual part), complex function expression $f(z)$ is an in-order traversal sequence by traversing complex function expression tree.

Both real, imaginary parts and the module of $f(z)$ can generate 3D images in MATLAB. Three images of $f(z)=\cos(z)*\log(-z^2)*\text{angle}(z^2)$ are shown as figure 1.

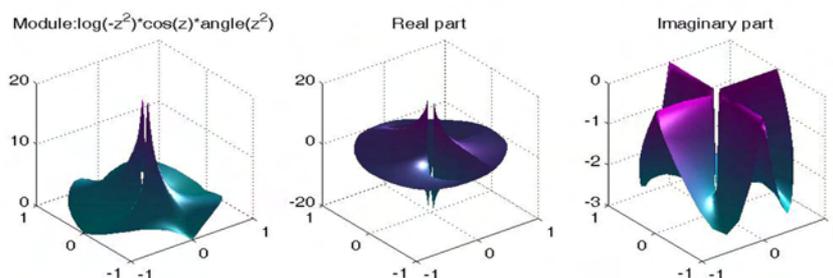


Fig. 1. Three images of $f(z)=\cos(z)*\log(-z^2)*\text{angle}(z^2)$

Next, we will present the performing process of the algorithm step by step.

Step 1: Initialize the population of chromosomes. The populations are generated by randomly selecting nodes in the set of operands and the set of operators to form complex function expressions. We use the stack to check whether such a complex function expression has properly balanced parentheses. Then, using parsing algorithm, the complex function expressions is read as a string of characters and the binary complex function expressions tree is constructed according to the rules of operator precedence.

Step 2: Get the fitness for each individual in the population via interaction with designer. The populations with high fitness will be shown in 3D form first. The designers can change the fitness value when they have seen the 3D images.

Step 3: Form a new population according to each individual's fitness.

Step 4: Perform crossover and mutation on the population.

Figure 2 shows two complex function expressions trees. Their expressions are $f(z)=\log(-z^2)*\cos(z)*\text{angle}(z^2)$ and $f(z)=\text{sqrt}(z)*\cos(-z^2)*\text{cot}(-z*0.5)$ respectively.

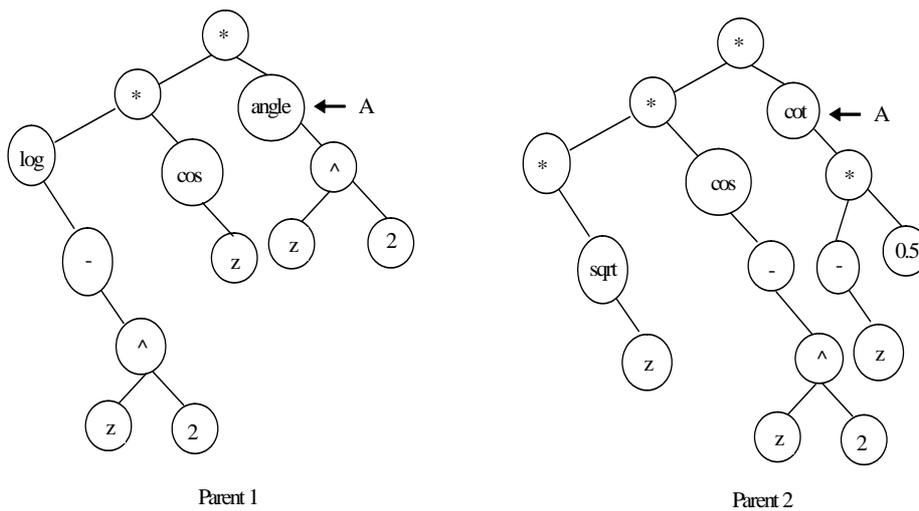


Fig. 2. Two parent trees with one crossover nodes

(1) Crossover operation

A crossover node is randomly selected in each parent tree. The sub-tree below this node on the first parent tree is then swapped with the sub-tree below the crossover node on the other parent, thus creating two new offspring. If the new tree can't pass the syntax check or its mathematical expression can't form a normal sketch shape, it will die.

Taking the two trees in figure 2 as parent, after the crossover operations by nodes 'A', we get a pair of children (figure 3).

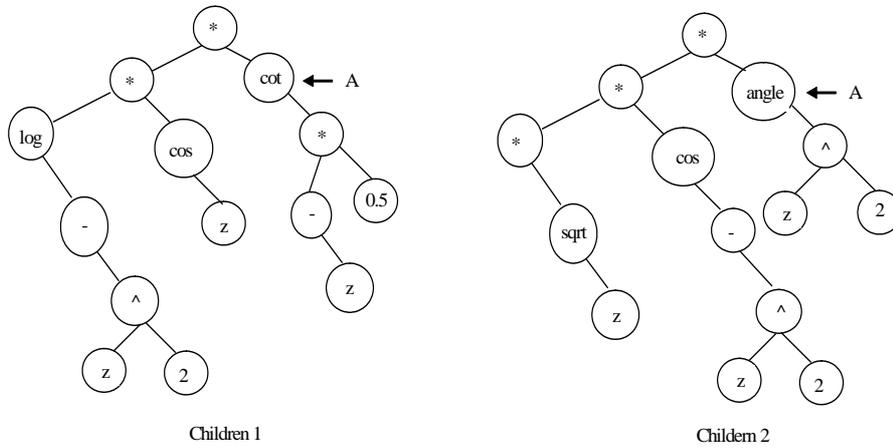


Fig. 3. The two children generated by a crossover operation

Figure 4 shows a group of generated 3D images by the module part of $f(z)$ correspond to figure 2 and figure 3.

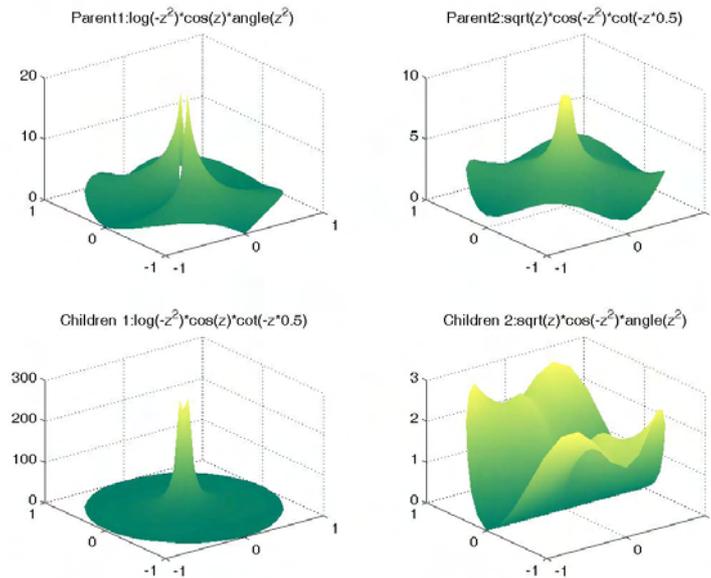


Fig. 4. The images correspond to the module of $f(z)$ in figure 2 and figure 3

(2) Mutation operation

The mutation operation works by selecting a random node in a single parent and removing the sub-tree below it. A randomly generated sub-tree then replaces the removed sub-tree. The offspring will die if it can't pass the syntax check or it can't form a normal shape.

Taking $f(z)=\log(-z^2)*\text{angle}(z^2)$ as a parent, one offspring generated by mutation operation is shown as figure 5. In which, child is generated by replacing node A and its sub-tree with new sub-tree. Figure 6 is the two images correspond to the real part of $f(z)$ in figure 5.

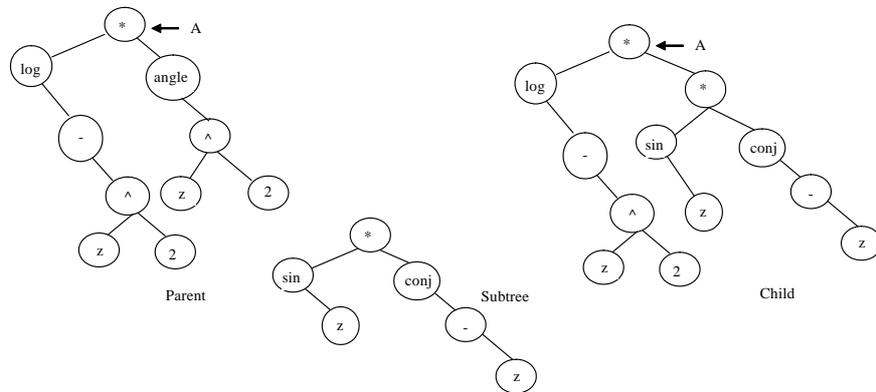


Fig. 5. One mutation operation

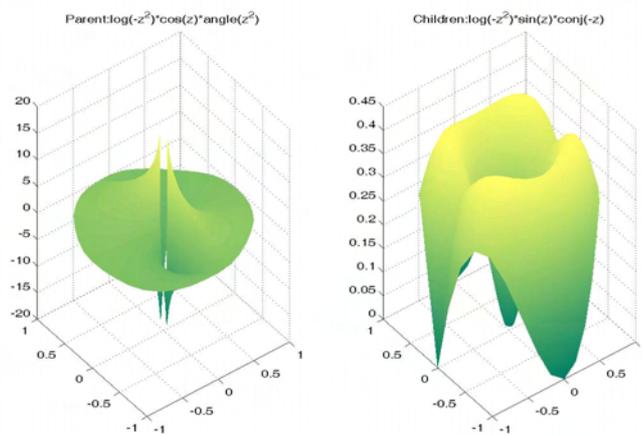


Fig. 6. The images correspond to the real part of $f(z)$ in figure 5

Step 5: If the procedure is not stopped by the designers, go to step 2.

This process of selection and crossover, with infrequent mutation, continues for several generations until it is stopped by the designers. This procedural design is integrated with a visualization interface, which allows designers to

interact and select from instances for design evolution. Then, the detail design will be done by designers with human wisdom [18].

The generated images are handled by designers using computer operations, such as rotating, cutting, lighting, coloring and so on. The interactive user interface can be seen in figure 7.

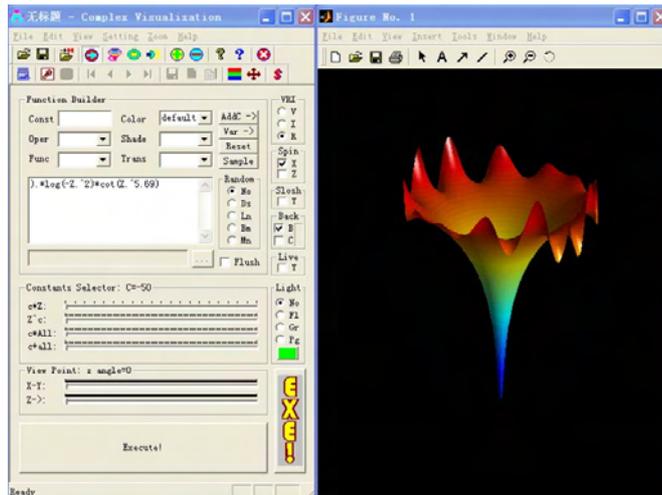


Fig. 7. The interactive user interface

6. Conclusions

This paper presents a novel computer-aided design system which uses tree-like expressions for representing complex mathematical functions and apply this tree structured genetic algorithm in generative design. Although looking simple, the framework employs a feasible and useful approach in a visual evolutionary computing environment. This system facilitates designers in two aspects: 1) diversify instances of design options; 2) enhance the possibility of discovering various potential design solutions.

Designing can be displayed as a dynamic and formal operation of an evolutionary procedure. This study employs the computer as an interface for generating a canonical population for selection. As for generative design, concepts of evolutionary selection are developed that explain different knowledge behaviours. The evolution of populations towards a stable state corresponding to the designers' consensus will be explored in our future work.

7. Acknowledgements

This project is founded by National Natural Science Foundation of China (No. 60970004, No. 60743010) and supported by Natural Science Foundation of Shandong Province (No. Z2008G02).

8. References

1. Innocent T. The language of iconica. In A. Dorin, J. McCormack (eds). *First Iteration: A Conference on Generative Systems in the Electronic Arts*, CEMA, Melbourne. 1999: 92-104.
2. McCormack, J., Dorin, A. and Innocent, T. Generative design: a paradigm for design research. In Redmond, J. et. al. (eds). *Proceedings of Future ground*, Design Research Society, Melbourne, 2004.
3. Bentley P..J. *Evolutionary design by computers*. Morgan Kaufmann Publishers, San Francisco, Calif. 1999.
4. Dorin, A. Aesthetic fitness and artificial evolution for the selection of imagery from the mythical infinite library. In Kelemen, J., Sosík, P. (eds). *Advances in Artificial Life, Proceedings of the 6th European Conference on Artificial Life*, vol. LNAI2159, Springer-Verlag, Prague, 2001: 659-668.
5. Frazer J. H. etc. *Generative and evolutionary techniques for building envelope design* .Generative Art 2002.
6. Gero, J. S., Kazakov V. An exploration-based evolutionary model of generative design process. *Microcomputers in Civil Engineering* 1996; 11:209-216.
7. Soddu, C. *New naturality: a generative approach to art and design*.Leonardo,2002,35(3):291-294.
8. Holland, J. H. *Adaptation in natural and artificial systems*. Cambridge, MS: MIT Press, 1975.
9. Sanghamitra, B. Sankar, K. P. *Classification and learning using genetic algorithms: applications in bioinformatics and web intelligence*. Berlin, New York : Springer, 2007.
10. Gero, J.S. and Peng, W. *Assisting Interactions in a Dynamic Design Process: A New Role for an Adaptive Design Tool*, in K Hampson, K Brown and P Scuderi (eds), *Clients Driving Construction Innovation: Mapping the Terrain*, CRC-CI, Brisbane, 2005: 201-210.
11. Soddu, C. *Generative art in visionary variations*, Art+Math=X conference, University of Colorado, Boulder, 2005.
12. Goldberg, D. E. *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley.
13. Koza, J. *Genetic programming, on the programming of computers by means of natural selection*, MIT Press, 1992.
14. Koza, J. *Human-competitive applications of genetic programming*. *Advances in evolutionary computing: theory and applications*. Springer-Verlag New York, Inc. , New York, NY, USA,2003.
15. William, B. L. Koza, J. R. *Genetic programming and data structures: genetic programming + data structures = automatic programming!*, Kluwer Academic Publishers, Norwell, MA, 1998.
16. Mckay, B., Willis, M. J., Barton, G. W. *Using a tree structured genetic algorithm to perform symbolic regression*, In *Proceedings of the First IEE/IEEE International*

Generative 3D Images in a visual evolutionary computing system

Conference on Genetic Algorithm in Engineering Systems: Innovations and Applications, Halifax Hall, University of Sheffield, UK. 1995, 487-498.

17. Liu, H., Tang, M. X., and Frazer, J. H. Supporting Evolution in a Multi-Agent Cooperative Design Environment. *Advances in Engineering Software*, 2002,33(6):319-328.
18. Liu, H., Tang, M. X., and Frazer, J. H. Supporting creative design in a visual evolutionary computing environment. *International Journal of Advances in Engineering Software*, 2004, 35(5): 261-271.

Biographical notes: Hong Liu is a Professor of computer science in the School of Information Science and Engineering, Shandong Normal University. She earned her PhD from Institute of Computing Technology, The Chinese Academy of Sciences in 1998. Her research areas of interest include evolutionary computing, computer-aided design, multi-agent system and generative design. She has published in journals such as *Computers & Industrial Engineering*, *Advances in Engineering Software*, *Applied Soft Computing*, *Journal of Computer-Aided Design and Computer Graphics* and *Journal of software*.

Received: July 27, 2009; Accepted: October 29, 2009.

Effective Semi-supervised Nonlinear Dimensionality Reduction for Wood Defects Recognition

Zhao Zhang¹ and Ning Ye^{1,2}

¹ School of Information Technology, Nanjing Forestry University,
210037 Nanjing, China
zzhang618@gmail.com

² School of Computer Science and Technology, Shandong University,
250100 Jinan, China
ye.ning@yahoo.com.cn

Abstract. Dimensionality reduction is an important preprocessing step in high-dimensional data analysis without losing intrinsic information. The problem of semi-supervised nonlinear dimensionality reduction called KNDR is considered for wood defects recognition. In this setting, domain knowledge in forms of pairs constraints are used to specify whether pairs of instances belong to the same class or different classes. KNDR can project the data onto a set of 'useful' features and preserve the structure of labeled and unlabeled data as well as the constraints defined in the embedding space, under which the projections of the original data can be effectively partitioned from each other. We demonstrate the practical usefulness of KNDR for data visualization and wood defects recognition through extensive experiments. Experimental results show it achieves similar or even higher performances than some existing methods.

Key words: semi-supervised learning, dimensionality reduction, wood defects recognition, (dis-)similar constraints.

1. Introduction

Many research and application areas need to deal with high-dimensional data, which leads to a hot of studying the methods of dimensionality reduction, whose aim is to find a meaningful low dimensional manifold from the original data. In many real applications, data lying in high-dimensional ambient space can be modeled by a low-dimensional nonlinear manifold.

Principal component analysis (PCA) [8], kernel principal component analysis (KPCA) [9] and kernel Fisher discriminate analysis (KFD) [12][14] are widely used in pattern recognition[15][17] [19]. Utilizing domain knowledge has been an important issue in data mining tasks [1]. In general, domain knowledge can be expressed in diverse forms, such as class labels and pairs constraints [2]. Semi-supervised dimensionality reduction is a new issue in semi-supervised learning, which learns from a combination of both labeled and

unlabeled data. In many practical applications, unlabeled samples are readily available but labeled ones are fairly expensive to obtain, so semi-supervised dimensionality reduction has attracted much attention.

Currently, automatic wood defects recognition is one of hot issues in the mechanical wood industry. Matti Niskanen [3] and P. Meinschmidt [6] used different technologies to study wood based defect detection problems. Here, we also present a solution the wood defects recognition and evaluate the performance of the proposed system by giving quantitative experiments. The block diagram of SVM classifier based detection system is shown in Fig.1.

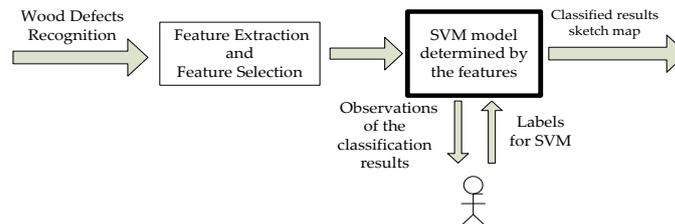


Fig.1. The diagram of SVM based recognition system.

Considering the pairs constraints can be derived from labeled data and can be automatically obtained without human intervention [1]. Here, we propose a semi-supervised nonlinear dimensionality reduction method (KNDR) with the pairs constraints for wood defects recognition. The rest of this paper is organized as follows. In section 2, we present a sensitivity analysis of KNDR and numerically evaluate it for wood defects recognition in section 3. We conclude this paper and raise some issues for future research in section 4.

2. Related Work

2.1. Linear dimensionality reduction method

For a set of wood samples together with some pairs of similar constraints and dissimilar constraints, we denote the domain knowledge containing similar and dissimilar pairs by S and D . For a set of data $X \in R^n$ and $X = (x_1, x_2, \dots, x_m)$, a simpler way of defining a criterion for the desired metric is to demand: if x_i and x_j are similar, pairs $\{x_i, x_j\}$ belong to set S , and D otherwise. To improve the tightness among similar pairs and separate dissimilar ones better, we consider shrinking distances between similar pairs, i.e. $(x_i, x_j) \in S$, by minimizing $\|x_i - x_j\|^2$, while expanding distances between dissimilar pairs, i.e. $(x_i, x_j) \in D$, by maximizing $\|x_i - x_j\|^2$. Given a set of multivariate data X with respect to the

pairs constraints, we aim to find a set of vectors $\omega = (\omega_1, \omega_2, \dots, \omega_d)$, such that the transformed low- dimensional representations $y_i = \omega^T x_i$ of x_i can preserve the structure of labeled and unlabeled samples as well as the pairs constraints, i.e. instances involved by S should be close while instances involved by D should be far as soon as possible.

Noticing that $\omega^T X$ means matrix ω basically projects the data on to a set of 'useful' features in embedding spaces [18]. Ideally, the set should be small that means the small rank for $\omega^T \omega$ or ω , as $\text{rank}(\omega^T \omega) = \text{rank}(\omega)$ is desired. Our aim is to minimize it by finding an eigen-decomposition $E(\omega^T \omega) = V \Lambda V^T$, then $\text{rank}(\omega^T \omega) = \text{rank}(\Lambda) = \|\Lambda\|_0$, but a direct minimization of this zero norm is very difficult and here we approximate it by the L2-norm $\|\Lambda\|_2 = \|\omega^T \omega\|_2$ in the model [18]. Then the $n \times d$ KNDR transformation matrix T_{KNDR} is defined in Eq.1, where $\|\cdot\|$ denotes the L2-norm.

$$T_{KNDR} = \arg \min_{\omega} \frac{1}{2} \|\omega^T \omega\|^2 + \frac{C_u}{2|N_u|} \sum_{(x_i, x_j)} \|\omega^T x_i - \omega^T x_j\|^2 + \frac{C_s}{2|N_s|} \sum_{(x_i, x_j) \in S} \|\omega^T x_i - \omega^T x_j\|^2 - \frac{C_d}{2|N_d|} \sum_{(x_i, x_j) \in D} \|\omega^T x_i - \omega^T x_j\|^2 \quad (1)$$

The two terms of Eq.1 expresses the average squared distance between all unlabeled samples in the embedding space and N_u is the number of unlabeled samples. $y_i \in R^d (1 \leq d \leq n)$ are the embedded data where d is the dimension of reduced space. N_s and N_d are respectively the numbers of samples under the similar and dissimilar constraints. The intuition behind Eq.1 is to let the average distances in the embedding space between instances involved by S be as small as possible, while distances between instances involved by D should be as large as possible. Since metrics between instances in the same class is typically smaller than those in different classes, here we add a scaling parameter C_s to balance the contribution of two terms and C_d for three terms in Eq.1. Intuitively, distances of samples involved in S should typically be close to the expected metric [20], so we empirically set $C_u = 1, C_s \geq 1$ and $C_d > 1$.

2.2. Kernelization

Implicitly in the kernel Hilbert space H connected to the kernel function K used. According to [4], a kernel is a function in the input space and at the same time is the inner product in the feature space through the kernel-induced nonlinear mapping. Since for each kernel there exist a mapping ϕ corresponds to a

scalar product and maps input patterns x_i to $\phi(x_i)$, here we define a mapping $\phi: R^n \rightarrow H^p (p > n)$ and choose *RBF* kernel for data projections:

$$K(x, x^T) = \exp^{-\left(\|x - x^T\|^2 / 2\sigma^2\right)} \quad (2)$$

Mika et al proved that every solution $\omega \in H$ (Kernel space) can be written as an expansion in terms of the mapped samples data [1]. i.e. the vectors ω^ϕ in the high-dimensional kernel space can be rewritten as

$$\omega^\phi = \sum_{i=1}^m \alpha_i \phi(x_i) = \phi(X) \alpha = K(X, \alpha) \quad (3)$$

where the input patterns in H are denoted by $\phi(X) = (\phi(x_1), \phi(x_2), \dots, \phi(x_n))$ and $\alpha = (\alpha_1, \dots, \alpha_d)$ are the transformation matrix and $x_i, i = 1, 2, \dots, m$ form a vector space. Especially, for all functions with the form of Eq.3, we get

$$\langle K(\cdot, X), K(\cdot, X^T) \rangle_R = \langle \phi(X), \phi(X^T) \rangle_H = K(X, X^T) \quad (4)$$

By substituting Eq.3 into Eq.1, we can update Eq.1 by

$$\begin{aligned} \min_{\alpha} J(\alpha) &= \frac{1}{2} \alpha^T (\phi(X))^T \phi(X) \alpha \alpha^T (\phi(X))^T \phi(X) \alpha + \\ &\quad \frac{1}{2} \sum_{i,j} \|\alpha^T (\phi(X_i))^T \phi(X_i) - \alpha^T (\phi(X_j))^T \phi(X_j)\|^2 R_{ij} \end{aligned} \quad (5)$$

where weights R_{ij} defined in Eq.5 satisfy the following formulations:

$$R_{ij} = \begin{cases} \frac{C_u}{|N_u|} + \frac{C_s}{|N_s|} & \text{if } \text{pairs}(x_i, x_j) \in S \\ \frac{C_u}{|N_u|} - \frac{C_D}{|N_D|} & \text{else if } \text{pairs}(x_i, x_j) \in D \\ \frac{C_u}{|N_u|} & \text{otherwise} \end{cases} \quad (6)$$

By substituting Eq.6 into Eq.5, we can rewrite Eq.5 as

$$\begin{aligned} J(\alpha) &= \alpha^T (\phi(X))^T \phi(X) \alpha \alpha^T (\phi(X))^T \phi(X) \alpha \\ &\quad + \frac{1}{2} \sum_{i,j} \|\alpha^T (\phi(X_i))^T \phi(X_i) - \alpha^T (\phi(X_j))^T \phi(X_j)\|^2 R_{ij} \\ &= \alpha^T (\phi(X))^T \phi(X) I (\phi(X))^T \phi(X) \alpha + \sum_i \alpha^T (\phi(X))^T \phi(x_i) W_{ii} (\phi(x_i))^T \phi(X) \alpha \\ &\quad - \alpha^T (\phi(X))^T \phi(X) R (\phi(X))^T \phi(X) \alpha \\ &= \alpha^T (\phi(X))^T \phi(X) (I + W - R) (\phi(X))^T \phi(X) \alpha \\ &= \alpha^T M \alpha \end{aligned} \quad (7)$$

where $M = KQK$, $K = (\phi(X))^T \phi(X)$ and W is a diagonal matrix whose entries are column (or row) sums of the matrix R , i.e. $W_{ii} = \sum_j R_{ij}$. $Q = I + W - R$ is called Laplacian matrix. The coefficient for the identity matrix is omitted here. Thus Eq.5 or Eq.7 can be simplified as

$$\text{minimize } J(\alpha) = \alpha^T M \alpha \quad \text{subject to } \alpha^T \alpha = I \quad (8)$$

Clearly, the problem in Eq.8 is a standard eigenvalue problem. Forming the Lagrangian of Eq.8 with the multiplier λ , i.e. $L(J, \lambda) = \alpha^T M \alpha - \lambda(\alpha^T \alpha - 1)$, by taking partial derivatives with respect to the variables α and zeroing it by $\nabla_{\alpha} L(J, \alpha) = 2M\alpha - 2\lambda\alpha = 0$, from which α can be easily solved by computing the eigenvectors of the matrix M corresponding to the d smallest eigenvalues.

3. Experiments and analysis

In the experiments, we investigate the performance of KNDR method for data visualization and wood defects recognition. We first take the Iris and Soybean datasets [16] for visualization. Then, we perform experiments on the wood image database [7,13] for wood defects recognition. Local Binary Pattern (LBP) [5,10] is used to extract the features from the wood images. In the experiments, support vector machine (SVM) classifier is used for classification. The process of semi-supervised learning can be described as follows: we first use the labeled samples to train a decision-making function, and then use the function to label the unlabeled samples. The tunable parameters C_u, C_S and C_D are always set to 1,1 and 20 respectively if without extra explanations.

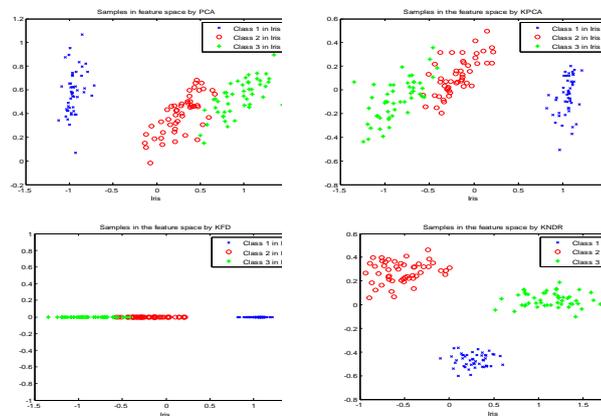


Fig. 2. Distributions of the features extracted by PCA, KPCA, KFD and KNDR on the Iris dataset ($D = 4, N = 150, T = 3, d = 2$)

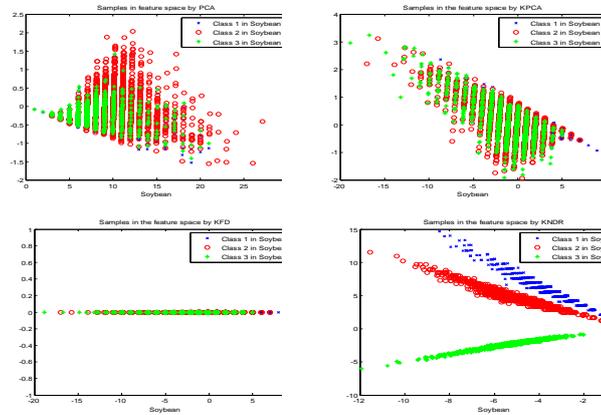


Fig. 3. Distributions of the features extracted by PCA, KPCA, KFD and KNDR on the Soybean dataset ($D = 8$, $N = 4177$, $T = 3$, $d = 2$)

3.1. Data visualization

In this subsection, we apply KNDR and some existing dimensionality reduction methods (i.e., PCA, KPCA and KFD) to the Iris and Soybean data sets with three classes and investigate how they behave in data visualization tasks. Figs.2 and 3 respectively depict the Iris and Soybean data embedded in the two-dimensional embedding space discovered by each method, where D is the dimension of the dataset, N is the number of instances, T is the number of classes and d is the number of selected features. The iris data set, popularly used for testing clustering and classification algorithms. The embedding spaces by PCA and KPCA are better, but both inferior to that of KNDR. The 'x'-class can be clearly observed by the four methods, however, for other two classes, PCA, KPCA and KFD tend to mix the data, but KNDR can separate them well. For the Soybean data set, the three classes are completely mixed in the original input space. PCA, KPCA and KFD can not work on the data set, while KNDR can keep in-class sample pairs close and between-class sample pairs apart effectively. Based on above experimental results, KNDR is found to be more appropriate for embedding the samples data than PCA, KPCA and KFD, implying that our primal goal has been successfully achieved.

3.2. Experimental results on wood defects recognition

There are many kinds of defects on the wood surfaces, such as discoloration, decadent, knots, etc. While knots can be seen more often and directly affect the quality of boards. Several varieties of defects are shown in Fig.4.

Effective Semi-supervised Nonlinear Dimensionality Reduction for Wood Defects Recognition

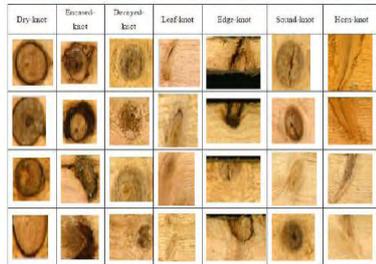


Fig. 4. The common kinds of knot defects.

Table 1. Sample class distribution.

Types	Number of samples		Total
	Training set	Test set	
Dry	69	27	96
Encased	19	4	43
Decayed	14	8	22
Leaf	37	3	70
Edge	62	1	63
Sound	126	10	186
Horn	32	5	37
Total	362	63	425

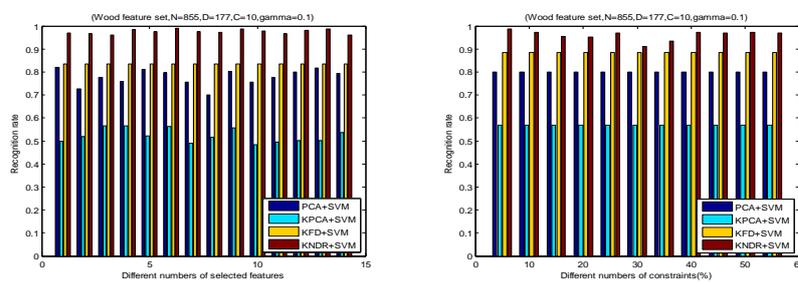


Fig. 5. Accuracy vs. different numbers of selected features and constraints.

For dimensionality reduction, we first transform the color wood images into three cues of R, G and B respectively and divide the images into many small blocks, and then extract the features from each block. Finally, we obtain 59 features from each cue, totally 177 dimensional features. In the experiments, the training set used in the experiments is very important. If the training set is chose improperly, the test results will be affected directly and make the system null. Here, we aggregately select 855 samples, including 417 positive samples (labeled 1) and 438 negative samples (labeled -1). We mainly take the knot and wood image database as basis, in which the sample class distribution is described in Table 1 and the number of different kinds of knot defects is also listed. The results reported here are based on a set of hundreds of wood images with over 200 labeled defects in the wood database.

Classification. Here, some experiments on the feature set are performed. The pairs constraints are obtained by randomly selecting pairs of instances from the whole data set, and creating similar or dissimilar constraints depending on whether the underlying classes of the two instances are the same or not. After obtaining the constraints, data without constraints in the whole data set are used as unlabeled data. Fig.5 displays the accuracy on the wood feature set under different numbers of selected features and constraints, where C is for SVM classifier and gamma is chose as the kernel parameters. From Fig.5, we find KDNR can almost always receive the highest recognition rate over 95% by comparing with the popular KFD, KPCA and PCA methods. The performance of KPCA is relatively poor in this feature set. With the increasing of the numbers of selected features, KNDR can keep stable for a

wide range and the performance of KFD method can also be thought better here.

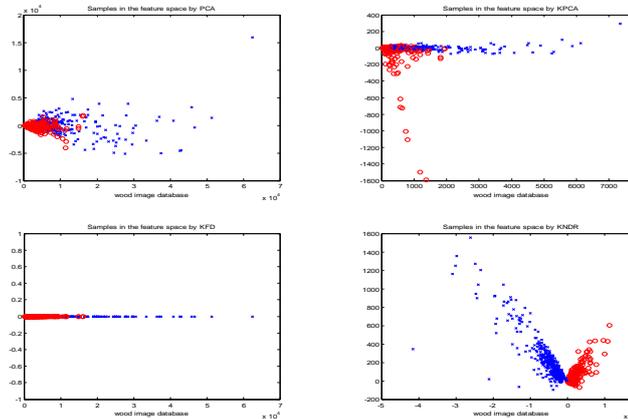


Fig.6. Distributions of the features extracted by PCA, KPCA, KFD and KNDR on the wood image database ($D = 177, N = 855, T = 2, d = 10$).

Fig.6 shows the distributions of the selected features extracted by PCA, KFD, KPCA and KNDR from the feature set. Intuitively, the embedded positive and negative samples are easier to be partitioned from each other in the feature spaces discovered by KNDR. Similar to section 3.1, PCA, KPCA and KFD tend to mix the projections of data of different classes.

Table 2. Results of the averaged accuracy and runtime.

Methods	Averaged accuracy (%)	Total runtime (s)
PCA+SVM	78.65	8.7
KFD+SVM	83.47	16.5
KPCA+SVM	55.80	20.6
KNDR+SVM	97.43	11.2

The above experiments have evaluated the learning ability of the proposed method. Next, the runtime performance of KNDR will be discussed. Table 2 gives the averaged accuracy and runtime of wood defects recognition under different number of selected features according to the left panel of Fig.5. From Table 2, the averaged accuracy of KNDR reaches 97.43% and the runtime performance is better than that of KPCA and KFD, but slightly worse than PCA, and we believe this is because PCA does use the linear projection.

Recognition results. There are no woods with same properties in color and surfaces show many varieties of texture characteristics, such as rough, fuzzy, etc. Even for the same species, the defects might greatly vary in shape and colors [11]. Furthermore, with large varieties of defects and the involvement of human factors caused the current detection methods are vulnerable to the

interferences on wood surfaces. Here, we barely emphasize particularly the category, but classify them as defects.

Let $A_i = [\psi_{i1}, \psi_{i2}, \dots, \psi_{in}]$, $A_j = [\psi_{j1}, \psi_{j2}, \dots, \psi_{jm}]$ represent two wood feature matrices, then the distance metric between them can be defined as

$$Dist(A_i, A_j) = \sum_{t=1}^n \|\psi_{it} - \psi_{jt}\|^2 \quad (9)$$

Supposing matrices A_1 and A_2 are use for storing the positive and negative samples, and are labeled by 1 and -1 respectively. For any new data point A , if $Dist(A, A_1) = \min Dist(A, A_j)$ and A_1 belongs to the negative samples with defects, then we judge A as a defect region, and normal region otherwise.

During the experiments, choosing an appropriate feature set is also very important. In the experiments, all the feature sets are selected by pixels 35*35 with good generalization ability. The number of constraints (60%) is randomly selected from the training set. Fig.7 displays the recognition results of the proposed method and the rectangular boxes are all located and drawn by the computers automatically.

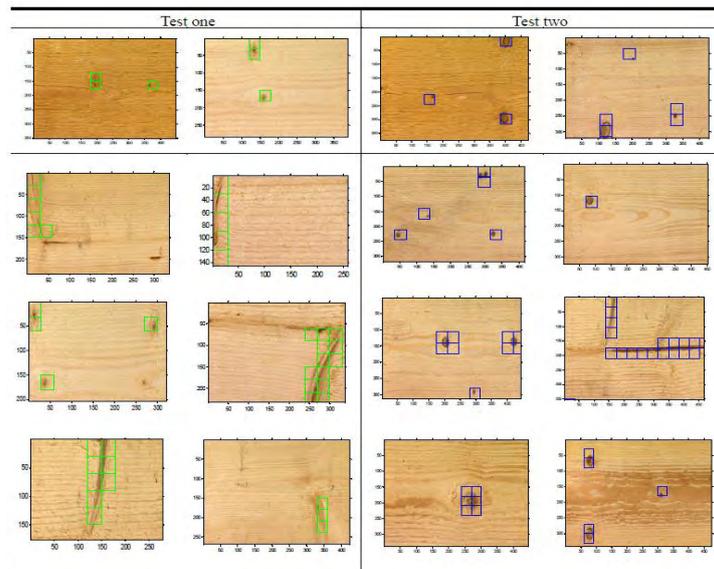


Fig.7. Experimental results of the wood defects recognition.

4. Conclusions and further works

The aim of this paper is to present a semi-supervised dimensionality reduction method called KNDR, focusing on domain knowledge in the form of the pairs constraints together with unlabeled data samples for defects recognition, not recognition results, however, the experimental results presented here are good or considerably better. We consider shrinking distances between similar pairs, while expanding distances between dissimilar ones in the embedding space. KNDR algorithm is interesting from a number of advantages: (1) KNDR is a standard eigenvalue problem and can be efficiently computed; (2) KNDR can preserve the structure of labeled and unlabeled samples and the constraints defined in embedding spaces; (3) For data visualization, the projections of the data in different classes can be effectively partitioned from each other; (4) In most cases, KNDR performs better than the classical PCA, KPCA and KFD methods. Furthermore, the runtime performance of KNDR is better than those of KPCA and KFD, but slightly worse than linear PCA.

Next, we will investigate how to choose the proper kernel parameters for the nonlinear model. Moreover, investigating whether KNDR can preserve the local structure of the original data is also an interesting future work.

5. Acknowledgments

This research was supported by the Innovation Foundation 2008 of Jiangsu Province of China under Grant No.164070265, the Innovation Program Foundation 2009 of Nanjing Forestry University, the Scientific Research Foundation 2009 of Jiangsu Province of China under Grant No.CX09S_013Z and the Natural Science Foundation of Jiangsu Province under Grant. BK2009393.

6. References

1. A.BAR-HILLEL, T. HERTZ, N. SHENTAL, AND D. WEINSHALL, Learning a mahalanobis metric from equivalence constraints, *Journal of Machine Learning Research*, 6, 937-965. (2005)
2. Xing, E.P., Ng, A.Y., Jordan, M.I., Russell, S.: Distance metric learning with application to clustering with side-information. In S. Becker, S.T., Obermayer, K., eds. In *Proceeding of NIPS*. MIT Press, Cambridge, MA, 15, 505-512. (2003)
3. Olli Silv'en, Matti Niskanen, Hannu Kauppinen. Wood inspection with non-supervised clustering. *Machine Vision and Applications*, 275-285. (2003)
4. Zhang, D.Q., Chen, S.C.: Clustering incomplete data using kernel-based fuzzy c-means algorithm. *Neural Processing Letters*, 18, 3, 155-162. (2003)
5. Ojala T, Pietikäinen M & Mäenpää T. Multi -resolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 7, 971-987. (2002)

6. P. Meinlschmidt, Wilhelm-Klauditz-Institut(WKI), Fraunhofer-Institute for wood research, Braunschweig. Thermo-graphic detection of defects in wood and wood-based materials. In Proceeding of the 14th international Symposium of nondestructive testing of wood, Hannover, Germany: (2005)
7. Sommardahl O, Usenius A. Wood samples image data. VTT Building Technology. (1999)
8. Hastie, T., Tibshirani, R., & Friedman, J. (Eds.). The elements of statistical learning: data mining, inference, and prediction. New York: Springer. (2001)
9. B. Schölkopf, S. Mika, A.J. Smola, G. Rätsch, and K. -R. Müller. Kernel PCA pattern reconstruction via approximate pre-images. In L. Niklasson, M. Bodén, and T. Ziemke, editors. In Proceeding of the 8th International Conference on Artificial Neural Networks, Berlin, Springer-Verlag, 147-152. (1998a)
10. Mäenpää T & Pietikäinen M. Texture analysis with local binary patterns. In: Chen CH & Wang PSP (eds) Handbook of Pattern Recognition & Computer Vision, 3rd ed, World Scientific, Singapore, in press (invited chapter).(2004)
11. Moring, I. and Silven, O. Automatic Visual Lumber Inspection: Experiments, Finnish Artificial Intelligence Symposium STeP-86, 2, 290-294. (1986)
12. Mika, S., Ratsch, G., Weston, J. et al.. Fisher discriminant analysis with kernels. In Y.H. Hu, J. Larsen, E. Wilson, and S. Douglas, editors, Neural Networks for Signal Processing, IX,41-48. (1999)
13. Olli Silven, Matti Niskanen, Hannu Kauppinen. Wood inspection with non-supervised clustering. Machine Vision and Applications, 275-285. (2003)
14. Mika, S., Ratsch, G., Müller, K. R., A mathematical programming approach to the Kernel Fisher algorithm. In: Leen T.K., Dietterich T.G., Tresp V. eds, Advances in Neural Information Processing Systems, Cambridge, MA:MIT Press, 13, 591-597. (2001)
15. J. Yang, A.F. Frangi, J.Y. Yang, D. Zhang, and Z. Jin. KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27, 2, 230-244. (2005)
16. C. Blake, E. Keogh, and C.J. Merz. UCI repository of machine learning databases. Department of Information and Computer Science, University of California, Irvine, 1998.[<http://www.ics.uci.edu/~mllearn/MLRepository.html>].
17. J.W. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. Face recognition using kernel direct discriminant analysis algorithms. IEEE Transactions on Neural Networks, 12, 117-126. (2003)
18. Ivor W. Tsang, James T. Kwok. Distance Metric Learning with Kernels. In Proceeding of International Conference on Artificial Neural Networks. (2003)
19. T. Xiong, J.P. Ye, Q. Li, V. Cherkassky, and R. Janardan. Efficient kernel discriminant analysis via QR decomposition. In Advances in Neural Information Processing Systems, 17. (2005)
20. Daoqiang Zhang, Zhi-Hua Zhou, and Songcan Chen. Semi-supervised dimensionality reduction. In Proceedings of the 7th SIAM International Conference on Data Mining (SDM'07), Minneapolis, MN, 629-634. (2007)

Zhao Zhang and Ning Ye

Zhao Zhang has graduated at the Faculty of Information Science and Technology in Nanjing Forestry University, in 2008. So far, he has authored several research papers. He is currently a master candidate for Computer Applied Science and Technology in Nanjing Forestry University. His research interests are machine learning and pattern recognition.

Ning Ye has graduated from Master's Degree from the Faculty of Test Measurement Technology and Instruments in Nanjing University of Aeronautics and Astronautics, in 1996. He got his Ph.D. from the Faculty of Computer Science and Technology in Southeast University, in 2006. He is currently a Professor of Computer Science in Nanjing Forestry University.

Received: May 05, 2009; Accepted: September 26, 2009.

MFI-Tree: An Effective Multi-feature Index Structure for Weighted Query Application

Yunfeng He and Junqing Yu¹

¹School of Computer Science & Technology,
Huazhong University of Science & Technology, 430074 Wuhan, China
yjqing@hust.edu.cn

Abstract. Multi-Feature Index Tree (MFI-Tree), a new indexing structure, is proposed to index multiple high-dimensional features of video data for video retrieval through example. MFI-Tree employs tree structure which is beneficial for the browsing application, and retrieves the last level cluster nodes in retrieval application to improve the performance. Aggressive Decided Distance for kNN (ADD-kNN) search algorithm is designed because it can effectively reduce the distance to prune the search space. Experimental results demonstrate that the MFI-Tree and ADD-kNN algorithm have the advantages over sequential scan in performance.

Keywords: Multi-Feature Index Tree; KNN; Aggressive Decided Distance for kNN; Video Retrieval.

1. Introduction

With the development of multimedia and network technologies, it is much easier to generate, access and manipulate video data than ever before. Facing the massive video data, traditional retrieval methods are not efficient by using only metadata of video, such as the name of video and the creator. People care more about the content of video data, so content-based video retrieval (CBVR) [1] is becoming an active research area in the area of video databases. There are two main methods for video retrieval in CBVR: one is semantic retrieval which matches the retrieval keywords with the semantic keywords extracted from video data; the other is sample retrieval which calculates the similarity distances between the features extracted from the sample, such as an image, a piece of video or audio etc, with the features of video data. Semantic retrieval method is simple and effective, but unfortunately, the existed technologies of CBVR still suffer from the semantic gap because computers can not directly "calculate" the semantic meaning from low-level features of video data, such as color, shape, texture, motion and audio information.

For sample retrieval method, it is not easy for people to find the right sample he wants. Normally, people browse the video database and select an image or a video clip as a sample to retrieve. Therefore, browsing is quite

important for sample retrieval method. In CBVR, video contents are usually described by multiple features, each of which is typically high-dimensional. For example, in MPEG-7 (Multimedia Content Description Interface) [2], a shot of video may be described by a 29-dimensional camera motion feature, and one keyframe of a shot may be described by a 12-dimensional color layout feature and a 31-dimensional homogeneous texture feature, etc. To support multi-feature queries, a high-dimensional index is needed to be built. Existed high-dimensional indexing technologies, such as M-tree [3] and VA-File [4], typically treat all different features homogeneously, which means the similarity distance is based on a static combination of feature weights. However, in sample retrieval, the weights of features are different for different people with different understanding to the sample. For example, an image sample is described by a color feature and a shape feature. Some people like its color and retrieve by using the weightages of (0.8, 0.2) for the color and the shape feature, while some people think the color feature is as the same important as the shape feature, and retrieve using the weightages of (0.5, 0.5). what's more, multiple high-dimensional features become ineffective with the dimension increasing.

In this paper, we propose a new indexing structure called Multi-Feature Index Tree (MFI-Tree) and a uniform similarity distance function is applied to ensure that the distance value of two objects is the one and only one in MFI-Tree building processing. MFI-Tree is a hierarchical tree structure which has two kinds of node, leaf node and cluster node. Leaf node represents a video data in the set, while cluster node represents an aggregate including some leaf nodes with a close distance. Division algorithm is important for the building and updating of MFI-Tree. Here, a new division algorithm which obtains several separate subsets is employed. To support K Nearest neighbors (kNN) queries, we propose a novel searching algorithm called ADD-kNN (Aggressive Decided Distance for kNN). To reduce the high-dimensional effect, ADD-kNN directly search cluster nodes in the last level of MFI-Tree. ADD-kNN is proved to be an efficient filter-and-refine approach which fast decreases the filtering value to avoid accessing data regions without objects belonging to the result-set.

The rest of the paper is organized as follows: related work is reviewed in section 2, while the structure of MFI-Tree and ADD-kNN searching algorithm is discussed in section 3 and section 4 respectively; in section 5, we make some experiments to evaluate the performance of the MFI-Tree and ADD-kNN algorithm; section 6 contains our conclusion and future work.

2. Related Works

The purpose of indexing is to improve the performance of queries. But for 30~50 dimension data, existed indexing techniques have failed to improve the performance of sequential scan due to the known "dimensionality curse" [5, 6]. To solve this problem, proposals in researches approximately belong to

three categories: dimensionality reduction, one-dimensional transformation, and data approximation [7].

Dimensionality reduction method maps the high-dimensional space into a low-dimensional space. The low-dimensional space is composed by some most important dimensions based on the correlation analysis of different dimensions, which is easy to be indexed by existed indexing techniques. For example, the dimension of color layout feature in MPEG-7 is reduced from 192 to 12. One-dimensional transformation includes Pyramid-Technique, iDistance, iMinMax, etc. The Pyramid-Technique [8] divides D-dimensional data space into 2-dimensional pyramid areas which share the center point of the space as a top and then cuts each pyramid area into slices, each of which forms a data page. Then the D-dimensional space is mapped to 1-dimensional space. iDistance method transforms a high-dimensional point into a 1-dimensional distance value with reference to its corresponding reference point [9]. iMinMax method maps points in high dimensional spaces to single dimension value determined by their maximum or minimum values among all dimensions [10]. One-dimensional transformation is efficient, however, because of the information loss in transformation, many candidates which are not results are calculated. In data approximation method, indexing is built on small and approximate representations which represent original data, such as VA-FILE (Vector Approximation File) [4]. The VA-FILE uses small vectors to represent the original data point and then sequentially scan the vector files to obtain candidates. However, the performance of VA-FILE is limited due to sequential scan. What's more, VA-FILE does not adapt to highly skewed data. Extended from VA-FILE, OVA-FILE uses the ordered approximation file where the approximations close to each other in data space are placed in the close positions based on VA-FILE [11].

With the development of multimedia database technology, there are some research works on multi-feature indexing structure. In [12], a single M-tree index is constructed for all the features, and principle component analysis and neural network is used to reduce dimension. But neural network training process is undesirable for very large data sets and M-tree structure is degraded in performance for dimensionality larger than 20. In [7], a multi-feature indexing structure using dimensionality reduction and B+-tree is proposed. Each feature is represented by two components: one is a 2-dimensional vector obtained by transforming each feature into minimum and maximum of a distance range, and the other is a vector of bit signatures which are set by analyzing each feature's descending energy histogram. This representation can effectively prune away points that are impossible to speed up query processing. However, B+-tree indexing is not suitable for browsing.

In fact, multi-feature indexing structures are given great attention with the development of video retrieval technology. But the existed researches put more emphasis on the indexing structure for the solution of "dimensionality curse" but less emphasis on the indexing system. The MFI-Tree structure and ADD-kNN searching algorithm proposed in this paper are suitable for retrieval and browsing application in video database.

3. MFI-Tree Structure

3.1. Uniform similarity distance function

In multi-feature video retrieval application, the similarity distance between two video objects is different for different weightages corresponding to different results. Set $F=(F_1, F_2, \dots, F_n)$ is a video data point described by n features, where F_i is the i^{th} feature and it is comprised of d_i dimensions. Thus, the similar distance function between video data point P and point O is following:

$$\begin{cases} Dist(P, O) = \sum_{i=1}^n w_i Dist_i(P, O) \\ \sum_{i=1}^n w_i = 1 \quad w_i > 0 \end{cases}$$

where $Dist_i(P, O)$ is the normalized distance value between point P and point O on the i^{th} feature, and w_i is the weight that describes the importance of the i^{th} feature.

For video data, different features describe different content of video, and apply different distance function. The distance of different feature has different value range. To generate a distance representing all features, each feature distance has to be normalized. We normalize $Dist_i(P, O)$ into the range of $[0,1]$ by the following normalization formula:

$$Dist_i(P, O) = \frac{Dist'_i(P, O)}{Dist_{i\max}}$$

Where $Dist_{i\max}$ is the maximal distance value of the i^{th} feature on two video data points, and $Dist'_i(P, O)$ is the distance value between point P and point O on the i^{th} feature. Here, the minimal distance value is not used because minimal distance value is easy to change when new object is inserted into the data set.

Note that the different weightages cause different distance value between P and O , but indexing building depends on the unique distance value between these two points. Therefore, a similar distance function for indexing building is applied. From normalization formula, it is easy to know that the distance value between two points is always less than the maximal distance value in the normalized distance value on n features. We can use the following formula to generalize the similar distance between two points:

$$Dist(P, O) = \max(Dist_i(P, O))$$

3.2. MFI-Tree Structure

MFI-Tree is a hierarchical tree structure which satisfies the need of video retrieval and browsing. Just like the M-tree, MFI-Tree has two kinds of node, leaf node and cluster node. Leaf node is used to save all video data points in video set, and figure 1 illustrates the structure of leaf node.

NodeID	FatherID	Feature F_1	...	Feature F_n	Dis
--------	----------	---------------	-----	---------------	-----

Fig. 1. The structure of leaf node in MFI-Tree

Where NodeID is the identification of current leaf node, and FatherID is the parent node ID of current leaf node, Feature F_i is the i^{th} feature value, and Dis is the distance value between current leaf node and the center object of the parent node.

Cluster node is used to save the cluster of some close leaf nodes. Different from the routing node of M-tree, the cluster node of MFI-tree is used not for retrieval, but for browsing to reduce the influence of the high-dimensional. The cluster node structure is shown in figure 2.

NodeID	FatherID	R	CenterID	BrowserID	Leaf-Num	isRoot
--------	----------	---	----------	-----------	----------	--------

Fig. 2. The structure of cluster node

Where NodeID and FatherID are used to build hierarchical tree structure for browsing, R is the covering radius of current node, namely the maximal distance of all the distances between each son node and the center node. CenterID and BrowserID is the identification of the center object and browsing object of current cluster node. LeafNum is the number of son node in current cluster node, and isRoot is the flag to show whether current node is the last level cluster node or not. When cluster node generates, a virtual point which does not really exist in video data set is used to be center object for decreasing the covering radius and the overlap area of cluster nodes. The browsing objects which present current cluster in browsing application can be saved to speed up the response time.

3.3. Building of the MFI-Tree

The building process of the MFI-Tree can be treated as a process that a large data set divides continuously into several small data sets, which needs three steps to accomplish.

Step 1: to calculate the maximal distance values of each feature. This is one of the main characteristics different from other dynamic indexing structures. For the distance value is normalized by maximal distance value of each feature, the changed maximal distance values of each feature may disable the MFI-Tree. That is why the MFI-Tree is called a semi-dynamic

indexing structure. The algorithm of finding the maximal distance values of each feature is shown in figure 3.

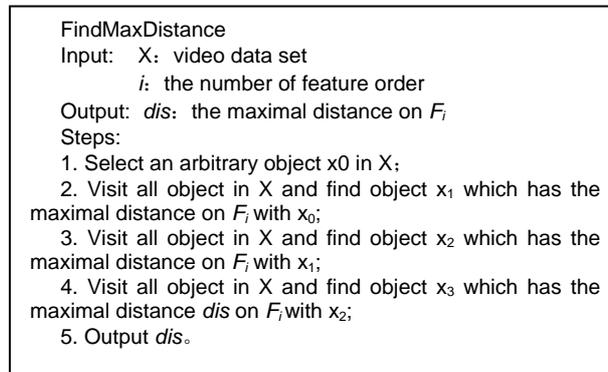


Fig. 3. Illustration of the maximal distance finding algorithm.

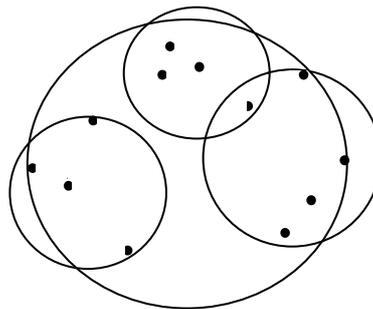


Fig. 4. Set division illustration in MFI-TREE.

Step 2: to divide data set. Considering browsing application, MFI-Tree does not divide data set into two subsets, but divide into several subsets based on the distribution of data sets. Figure 4 gives the illustration of the set division process. Firstly, we find the farthest pair of object A and B in data set using the algorithm like finding the maximal distance algorithm, and then object A and B are inserted into two new subsets separately. Secondly, calculate the minimal distance value between objects which is not distributed with the first object of each subset. The minimal distance value is denoted to d_{min} . If d_{min} is less than threshold value AddDis, it means the object is close enough to the corresponding subset and is inserted into it, such as object K, L, H and G in figure 4. If d_{min} is more than threshold value NewSetDis, it means the object is far away from all subsets, such as object C in figure 4, then a new subset is created and the object is inserted into the new subset. If d_{min} is more than AddDis and less than NewSetDis, the object does not execute insert processing, such as the object D, I and J in figure 4. Finally, calculate the minimal distance value between objects which is not distributed with the first object of each subset, and the object is inserted into the

corresponding subset. The algorithm of dividing the data set is shown in figure 5.

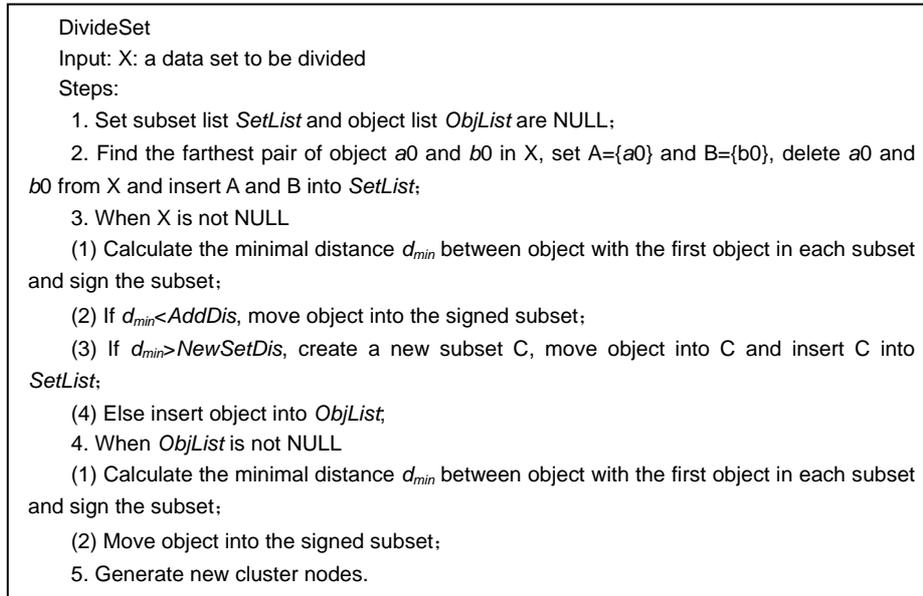


Fig. 5. Set division algorithm in MFI-Tree

After all objects in data set are inserted, all subset nodes are generated. For each subset, the center point of the farthest pair of objects is the center object of the subset, the maximal distance value between objects in the subset and the center object is the covering radius of the subset, and the object which has minimal distance to the center object is the browsing object of the cluster.

In the division algorithm, two important threshold values are used. *NewSetDis* is used for creating a new subset and is calculated by the following formula:

$$NewSetDis = \delta \cdot D_{max}$$

where D_{max} is the maximal distance between two objects in the data set to be divided, and the δ is the degree of the division. Normally, new subsets are created when δ is more than 0.5. But for the distance function of feature is not Euclid distance and the distance value is the maximal distance value, δ is usually from 0.6 to 0.8 due to reducing the number of subsets. 0.7 is used in our experiments. *AddDis* is used for the inserting operation and is normally equal to half of the *NewSetDis*.

Step 3: to check new subset. If the subset meets the conditions of division, the subset can execute division, else stop. There are two conditions for set division. One is *LeafNum*, which is the number of the objects in set. The bigger *LeafNum* increases the calculation, but the smaller *LeafNum* increases the number of cluster nodes and reading times. The other is the covering

radius R . The bigger R decreases the performance of pruning, while the smaller R increases the number of cluster node. Normally, LeafNum is depending on the organization of data. And the maximal R value is different for different application. In multi-feature indexing structure, most data points do not cluster into a small area because of using the maximal feature distance as object distance. The maximal of R is from 0.1 to 0.3, and in our experiments, this value is set to 0.3.

3.4. Insertion and Deletion Algorithm

When the MFI-Tree is generated, the insertion and deletion operations can be enabled if the maximal distance values of each feature are not increased.

In the process of inserting, the first is to find all cluster nodes in the last level of MFI-Tree, and to calculate the minimal distance between the insert object and the cluster node. The second is to insert object into corresponding cluster node and update the LeafNum and R of the ancestor nodes up to top. Finally, if the cluster node to be inserted meets the conditions of division, this cluster node executes division operation.

The deletion operation is similar to the insertion. The first step is to find and delete object from MFI-Tree, and then update the LeafNum value of ancestor nodes up to top. Lastly, if the LeafNum value of the cluster node is smaller than the conditions of division, the father node of this cluster node will delete all its son nodes, and then execute division operation.

4. ADD-kNN Searching Algorithm

Ordinarily, video content similarity queries contain some elementary types, such as Range Query, k nearest neighbor query (kNN query), etc. Using a range search, it is difficult to specify a maximal distance as the constraint without some knowledge of the data and distance function. An alternative way is to use kNN query which finds k nearest neighbors to the given query object.

There are two ways of approaching kNN query, range query and filtering. For the range query, query data set uses a given distance R . If the number of candidates is more than k , then k nearest neighbors are returned. Else, continually query data set uses increased R value until k results is found. Obviously, the performance of the mentioned methods relies on the value of R , the bigger or the smaller of the value can decrease the query's efficiency. The distance functions of high-dimensional features have the property of triangle inequality, that is:

$$\forall x, y, z \in D, d(x, z) \leq d(x, y) + d(y, z)$$

Filtering method uses triangle inequality to prune away a lot of impossible objects, and then calculate the distance between the query object and

candidates. This method can improve the performance by reducing the unnecessary calculation.

Comparing with the traditional indexing structure, MFI-TREE structure has an important characteristic. In the building processing of MFI-Tree, the distance between two objects is the maximal distance value in the normalized distance value on n features. But in query processing, the distance between two objects is calculated by given weightages. That is to say, the covering radius of cluster nodes in MFI-TREE is bigger than the real radius of cluster nodes calculated by given weightages in most cases. Because the covering radius is an important parameter for filtering methods, some useless cluster nodes cannot be pruned away easily by using bigger radius in query processing.

To avoid performance decreasing, a new kNN search algorithm, called Aggressive Decided Distance (ADD-kNN) is proposed. The main idea of ADD-kNN algorithm is to fast decrease the filtering value, and to effectively filter most of unnecessary objects. The ADD-kNN search algorithm can be characterized by using the following steps:

Step 1: create the query object according to the sample, features and its weightages.

Step 2: empty the query node list and result list, and set value D for filtering to zero.

Step 3: different from breadth-first search and depth-first search algorithms, our method traverses all the cluster node in the last level of MFI-Tree, that are all the cluster node whose value $isRoot$ is equal to 1.

Firstly, the real distance value d with the given weightages is calculated between the query object and the center object of the current cluster node. Then, the minimal distance value D_{min} is calculated by using the following formula:

$$D_{min} = d - R$$

where R is the covering radius of the current node.

When value D_{min} is less than zero, the query object is in the current cluster, and the objects in cluster node are all candidates. The objects in cluster are inserted into result list one by one according to the real distance to the query object. If the number of the objects in result list is more than k and the real distance is more than filtering value D , the object is not inserted. After insertion operates, if the number of the objects is more than k , the filtering value D is equal to the k_{th} smallest distance value in the result list, and then the object whose distance value is more than D is removed from result list. If the number of objects is less than k , the filtering value D is equal to the biggest distance value of objects in result list.

When value D_{min} is more than zero, the query object is outside the current cluster, and the cluster node is inserted into query node list.

Step 4: when the query node list is not empty, traverse all nodes in query node list. If the number of objects in result list is more than k and value D_{min} of the node is more than the filtering value D , the node is removed from query node list. Else, the objects in the node will execute insertion like the step 3, and then that can be removed from query node list.

Step 5: output the result list.

According to the characteristics of MFI-Tree structure, ADD-kNN research algorithm calculates the real similar distance values which are necessary for reducing the filtering value, decreasing the spending of sorting and improving the performance of retrieval.

5. Experiments

The 12-dimensional color layout feature and the 80-dimensional edge histogram feature in MPEG-7 are used in our experiments because these features are compact and effective to describe image contents. The data set for experiments were conducted by using 400 pieces of video clips from 30 movies, including various movie types like action, comic, comedy and science fiction, etc. The color layout and edge histogram feature are extracted from 10000~30000 images, which are the keyframes derived from the movie clips per 20 frame, using the extraction algorithm and distance function mentioned in MPEG-7. All the experiments were performed on an IBM T61 portable computer. The operating system is windows XP and the database is Oracle9i.

5.1. Effect of data size

We generate MFI-TREE structures and M-tree structures for 10000, 20000 and 30000 keyframes respectively, and then use 6 images derived from test video clip as query objects to execute k-NN query, where k is equal to 20. The experimental results are illustrated in figure 6 (a) and (d). It can be easily found that the average similar distance calculating times of the MFI-Tree structure using ADD-kNN algorithm are less than the calculating times of M-Tree structure. The reason is that the covering radius of current node in M-tree structure which uses uniform similarity distance function is bigger, the k-nn research algorithm of M-tree cannot prune away most nodes, and accessing hierarchic structure of the M-tree structure cause its inefficiency.

5.2. Effect of weighted queries

We use 6 images derived from test video clips as query objects to execute k-NN query with different weightages on the set of 20000 keyframes. Figure 6 (b) and (e) shows the average retrieval time and the average distance calculating times when the weight of colorlayout feature changes from 0.1 to 0.9 while the sum of two feature weights is equal to 1. It is shown that the average retrieval time is the highest when the weight of colorlayout feature is 0.4, and the average distance calculating times is the highest when the weight of colorlayout feature is 0.3, for the retrieval time is decided by distance calculating times and database accessing times.

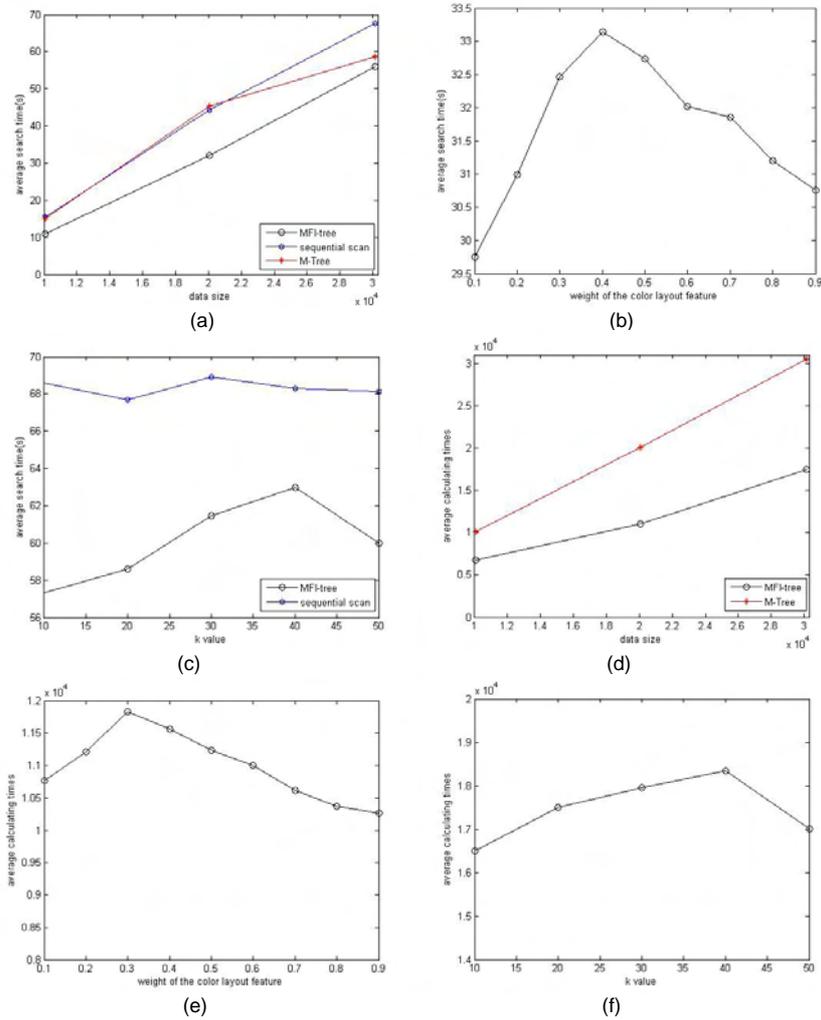


Fig. 6. Experimental results

5.3. Effect of k value

We executed kNN queries when the value of k changes from 10 to 50 on the set of 30000 keyframes. The results are illustrated in figure 6 (c) and (f). As we can see, ADD-kNN achieves the worst performance when k value is equal to 40. Actually, the maximal LeafNum value of the cluster nodes in indexing structure definitely influences query performance. Biggish value may lead to biggish covering radius, while less value may lead to more cluster nodes. So,

the maximal LeafNum value is decided by the size of data set and the k value which is usually used in applications.

6. Conclusion

In this paper, we have proposed a multi-feature indexing structure—MFI-Tree and ADD-kNN search algorithm for the weighted query application of video retrieval. Based on uniform similarity distance function, MFI-Tree is built to index multi-features of video data. MFI-Tree is a hierarchical structure which can be used for browsing effectively. The ADD-kNN search algorithm directly search the last level cluster nodes in MFI-Tree structure to reduce the high-dimensional effect of the indexing structure, and fast minimize the filtering value to prune most unnecessary data away. The experimental results demonstrate that the MFI-TREE and ADD-kNN search algorithm are effective and efficient for weighted query application.

However, in the division operation of MFI-Tree, the overlap areas of the subsets still exist, which can influence the performance of retrieval. What's more, the features to describe the video content are not easy to be understood by people, so how to create an effective interface for weighted queries application will be one of our future work.

7. Acknowledgement

This paper is financially supported by the National Natural Science Foundation of China under Grant No.60703049; the “Chen Guang” Foundation for Young Scientists of Wuhan under Grant No. 200850731353.

8. References

1. Michael S. Lew, Nicu Sebe, Chabane Djeraba, and Ramesh Jain: Content-based multimedia information retrieval State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications and Applications*, 22(1), 1–19 (2006)
2. ISO/IEC/JTC1/SC29/WG11 (MPEG). Text of ISO/IEC 15938-3 Multimedia Content Description Interface – Part 3: Visual. Final Committee Draft. Singapore, (2001)
3. Paolo Ciaccia, Marco patella, Pavel Zezula. M-tree: An efficient Access method for Similarity Search in Metric Spaces. In proceedings of the 23rd VLDB conference, Athens, Greece, 426-435 (1997)
4. R. Weber, H.J. Schek, and S. Blott: A Quantitative Analysis and Performance Study for Similarity Search Methods in high-Dimensional Spaces. In proceedings of the 24th VLDB Conference, New York, USA, 194-205 (1998)

5. Gisliir. Hjaltason and Hanan Samet: Index-Driven Similarity Search in Metric Spaces. *ACM Transactions on Database System*, 28(4), 517-580 (2003)
6. Christian Bohm, Stefan Berchtold and Daniel A.Keim: Searching in High-Dimensional Spaces-Index Structures for Improving the Performance of Multimedia Databases. *ACM Computing Surveys*, 33(3), 322-373 (2001)
7. H.V. Jagadish, Beng Chin Ooi, Heng Tao Shen, and Kian-Lee Tan: Toward Efficient Multifeature Query Processing. *IEEE Transactions On Knowledge and Data Engineering*, 18(3): 350-362 (2006)
8. S. Berchtold, C. Bohm, H.P. Kriegel: The Pyramid-Technique: Towards Breaking the Cures of Dimensionality. In proceedings of Int. Conference On Management of Data, ACM SIGMOD, Seattle, Washington, 142-153 (1998)
9. C. Yu, B.C.Ooi, K.L.Tan, H.V.Jagadish: Indexing the Distance: An Efficient Method to KNN Processing. In proceedings of the 27th VLDB Conference, Roma, Italy, 166-174 (2001)
10. B.C. Ooi, K.L. Tan. C. Yu, S.Bressan: Indexing the Edges-A simple and Yet Efficient Approach to High-Dimensional Indexing. In ACM SIGMOD 19th Symposium on Principles of Database Systems (PODS), Dallas, Texas, 166-174 (2000)
11. Hong Lu, Beng Chin Ooi, Heng Tao Shen, and Xiangyang Xue: Hierarchical indexing structure for efficient similarity search in video retrieval. *IEEE Transactions on Knowledge and data engineering*, 18(11), 1544-1559 (2006)
12. A.H. Ngu, Q. Sheng, D. Huynh, and R. Lei: Combining Multivisual Features for Efficient Indexing in a Large Image Database. *VLDB*. 9(4): 279-293 (2001)

Yunfeng He is currently a PhD. Candidate and a member of faculty at HuaZhong University of Science and Technology in China. His research interest focuses on video data index and retrieval.

Junqing Yu is the corresponding author of this paper. He received his PhD in Computer Science from Wuhan University in 2002. He is currently an Associate Professor at the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China. His research interests include digital media processing and retrieval, multi-core programming environment.

Received: May 14, 2009; Accepted: August 07, 2009.

A Content-based Dynamic Load-Balancing Algorithm for Heterogeneous Web Server Cluster

Zhang Lin¹, Li Xiao-ping², and Su Yuan²

¹School of Electricity and Information Engineering, BUCEA, Beijing, 100044, China
zzcmeng0806@sina.com

² Department of Computer Science and Technology, Beijing Institute of Technology,
Beijing 100081
lxpmx@x263.net

Abstract. According to the different requests of Web and the heterogeneity of Web server, the paper presents a content-based load-balancing algorithm. The mechanism of this algorithm is that a corresponding request is allocated to the server with the lowest load according to the degree of effects on the server and a combination of load state of server. Besides, apply a method of random distributing base-probability to assign each request to an appropriate server in terms of their weight. All the parameters that will be used in the algorithm can be acquired by simulated test. Experimental results suggest that this algorithm can balance the load of web server clusters effectively, make full use of the existing source of software and hardware, highly improve the server's performance, and even make the best use of the web server.

Keywords: Web server cluster; load-balancing; dynamic feedback

1. Introduction

A Web cluster refers to a Web site that uses two or more server machines housed together in a single location to handle user requests. Although a large cluster may consist of dozens of Web servers and back-end servers, it uses one hostname to provide a single interface for users^[2].

Request distribution and load balance are key technological means to realize web server clusters, but most existing request distribution algorithms are static ones and they don't have any optimization strategy and don't synthetically consider the dynamic parameters such as the utilization ratio of CPU and memory. With the apply of Jsp、 database and multimedia, the request coming from different client will bring different load on web server. So those existing equilibrium algorithms are inefficient and erratic and don't fit to construct heterogeneous Web server clusters. This paper considers the heterogeneous characteristic of hardware and software on every node and puts forward a content-based load-balancing algorithm of application layer—

QSC-Load Balancing algorithm. Experimental results suggest that it can improve the response time and throughput of the system.

2. Analysis of the content-based request distribution strategy

In the cluster system based on TCP/IP request distribution^[1], the fore-end is transmitted based on the IP and the target port, without knowing the clients' submission. However the distribution based on the application layer means the target content must be informed beforehand (for instance the URL information carried by the message head of HTTP). In addition, information will be configured and the request will be distributed to the corresponding back-end server's node according to the matching rules. Take web service as an example, it requests the client and the fore-end to build up a complete TCP connection so as to acquire the head message of HTTP. After receiving the HTTP requirement, the fore-end starts to analyze and apply the suitable distribution strategy in order to transfer the request to the back-end server by which the request will be carried out^[6].

As for the content, different distribution strategies have different definitions^[3]. Rough divisions can be made according to the types of the document, such as HTML, GIF, and CGI. Relatively detailed contents mean the complete parts of URL in the HTTP's head message, that is the web object's oriented message expected by the client. This paper discusses the latter content, which is the specific URL message^[7].

HTTP's text is divided into two parts: the first part describes the format and content of the request message sent by client; the second part describes the format and content of the answer message sent by Web server. This paper is only concerned with the contents of the request message. HTTP request consists of two parts, one is a request head with line structure and the other is request body. The request head includes several parts as follow:

1. Method: located in the first line of request head, pointing out the request operational types. There are four main methods: GET、PUT、 POST and HEAD.

2. URL: located in the first line of request head and after the method, pointing out the location of the target and other information.

3. Protocol version: pointing out the HTTP protocol version number applied by the client browser.

4. Compressed mode: pointing out the data compressed mode supported by the client browser.

In the HTTP request message, the target URL after the GET method in the first line is an important gist applied in the request distribution system.

3. Design of QSC-load balancing algorithm

3.1. Theoretical analysis of QSC-loading balancing

From the logical resource point of view, the web server can be regarded as one composed of several resource nodes providing all kinds of service. It is an important problem of how to share the entire service ability in the entire server, without overloading a single node^[5].

According to the degree of influence of the server, this paper divides the request into static request, database operating request and safety request. Each request is endowed with specific weight-value. The calculation formula of request weight-value is defined as follow:

$$w[i] = \text{CPU process time} * a + \text{memory use} * p + \text{disk access time} * (1-a-p)$$

$w[i]$:relative weight-value of request i ; a , p : relative weight-value coefficient, namely the importance of every class resource.

In order to simplify the problem, this paper doesn't consider the practical factors, such as the choke point and the wastage in the internet transmission. Fro the convenience of the research, time is equably divided into discrete time slices(0 T) (T 2T)...(kT (k+1)T)....

Definition 1: the start moment of every time slice indicates this period of time,such as kT represents $(kT (k+1)T)$.

Definition 2: C indicates the processing ability of the server in the unit time.

Definition 3: $C(kT)$ indicates the processing ability of a web server within kT .

Definition 4: $S(i,kT)$ indicates the needed processing ability of mission i within kT .

Definition 5: $N(kT)$ indicates the allowed number of mission within kT .

So we can get an equation as follow:

$$C(kT) = c * T \geq \sum_{i=1}^{N(kT)} S(i, kT) \quad (3-1)$$

When the result of equation (3-1) is true, the request i is allowed to enter the service queue, otherwise it will be refused. This equation shows that the maximum of the operating lag of the request is T and the value of T affects the system efficiency. So we must choose the different T according to the processing ability of the system.

3.2. Description of QSC-load balancing

3.2.1 Performance parameter of server

In order to achieve higher system throughput and shorten the client's feedback time, the algorithm adopted both static and dynamic parameters to reflect the capability of sever^[4].

1. Static performance parameter

The parameter of software and hardware of the web server determine its capability. During the processing of the web server cluster, these parameters are changeless, so they are also called static performance parameters. This paper selected the following static performance parameters:

(1) CPU processing ability: in the cluster system, the process of sever monitor inspects the number and type of CPU, and endows, according to previous experience value, the server with a moderate CPU processing ability value.

(2) Memory parameter: the process of sever monitor inspects the size of computer's physical memory and virtual memory, so as to get the memory parameter.

2. Dynamic performance parameter

During the processing of cluster system, each web server's load is changing as time goes on, the system has to estimate the load-balance according to the real-time server load, and these are called dynamic performance parameters. This paper selected the following dynamic performance parameters:

(1) Processor utilization ratio: it can reflect the busyness degree. The process of sever monitor inspects the CPU utilization ratio termly, so as to confirm the CPU's load.

(2)Memory utilization ratio: the size of the server memory changes as the system runs. The process of sever monitor inspects the utilization ration of physical memory termly, so as to confirm the server memory's load.

(3)Network flow: the network data are mainly transferred through TCP mode in cluster system, the process of sever monitor inspects the packages received and sent by server termly, so as to get the server's load.

3.2.2 Division of request types

1. Issuance type: mainly providing static information. This request is the easiest one, including html or other undivided documents.

2. Affair type: mainly coming from the operation on dynamic database, and the condition of operation is provided by user via a dynamic HTML page. Because it needs dense access to the disk, and the reserved operation to obtain the type of the document, the weight-value is large.

3. Dynamic page type: needs to analyze page information dynamically, including obtaining the extended name such as jsp, asp and php, the weight-value is large.

4. Security type: provides static, dynamic and security information transmissions. During the security transmission, encrypting operate will consume large numbers of CPU resource and it needs dense access to the disk, so the weight-value is large.

5. Multi-media type: provides with real-time audio and video services. The weight-value is large.

3.2.3 Description of QSC-load balancing

When the linking number N exceeds the prescriptive limit value N_{max} , which is determined by the server ability, it will get a view. Here the processing time becomes longer and we call it key state, N_{max} as key number. In order to insure the server performance, N should avoid approaching N_{max} . We can confirm that the server is loaded by the web linking number exceeding the limit value.

In web cluster system, Cluster_Server_Table has requests of n style and servers of m brand, $1 \leq i \leq N$, each request is endowed with given value $w[i]$ via Service_Mapping_Table(SMT), $1 \leq i \leq N$, so the request load of server j is:

$$Load[j] = \sum_{i=1}^N R(i)[j] * w[i] \quad (3-2)$$

$R(i)[j]$ is $R(i)$ requests to connect with server j , so the request load of server j is:

$$Load[j] = \sum_{i=1}^N R(i)[j] * w[i] + w[k] \quad (3-3)$$

Server j has disposed request $R(i)$, so the request load of server j is:

$$Load[j] = Load[j] - w[i] \quad (3-4)$$

The judging measure of load-balancing:

Judging method of load-balancing: read in every line of CST, and then write into a record array, thus the array's every record represents all information of one server. Take the first working server's information (assume server M) as benchmark and put it and other working server's (assume server i) information into equation (3-5) to compare:

$$ratio = \alpha \times \frac{\lambda_1 \times C_i}{\lambda_M \times C_M} + \beta \times \frac{\lambda_2 \times M_i}{\lambda_M \times M_M} + \gamma \times \frac{N_i}{N_M} + \theta \times \frac{R_i}{R_M} \quad (3-5)$$

Parameters' meaning are as follow:

- $\lambda 1$: CPU processing ability.
- $\lambda 2$: memory parameter.
- C: CPU utilizing ratio.
- M: memory use ratio.
- N: network flux.
- R: request load.
- α : CPU comparative weight-value.
- β : memory comparative weight-value.
- γ : network comparative weight-value.
- θ : request load comparative weight-value.

The initial values of α 、 β 、 γ 、 θ are all 1. According to the cluster running fact we can increase or decrease some weight-value to enhance or weaken the corresponding load performance. The static and dynamic parameters are all considered when we are comparing the performance of CPU, memory and network.

We will have two problems when calculating the lightest load server in the CST. On the one hand, constantly calculating the computer's load will result in reduction of the computer's capability. On the other hand, when two continuous overloads are distributed into the same sever, it may respond quite slowly to the client's. In order to solve those problems, this paper applies a method of random distributing base-probability to assign each request to an appropriate server in terms of their weight. First of all, calculate the comparative weight-value to ensure its probability space via each sever, all the cluster server system space is 1. Fig.1 shows 4 probability spaces of the servers with the relatively weight-value of 0.1, 0.15, 0.25 and 05.

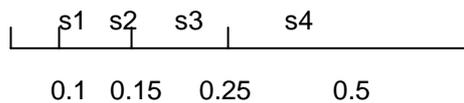


Fig. 1. Probability space of servers

When a new request mission reaches distribution mechanism, it temporarily calculates a random number between [0,1].According to the random number's placement in the probability space, make sure the target server of this transmission. Because the lower weight-value server has larger probability space, it will have higher allocation probability, meanwhile, each transmission is an individual event that can neither be affected by the former transmission, nor be affected by the next transmission operation, so it can reach more average effects. This is the improved algorithm of QSC-load balancing. Here is the description:

```
Load synchronization ( ) {
  For each server j in CST
```

```

    if request R(i) responses
        set the request load of server j: load[j]'=load[j]- w[i];
    if delay service queue is not null
        Pop the first request in the queue;
        distribute the out request according to request-distribution();
    Heart beat (); //inspect current load and the server's useableness in CST
} //Load synchronization ()
Request distribution (){
    if request r arriving
        Search Type(); //search in Service_Maping_Table(SMT) and identify the
        type of request r
        calculate the load of server j:load[j]'=load[j]+ w[i];

    if the number of server exceeds connection number  $N_{max}$ 
        set the load of server j:load[j]'= infinite;
    else
        For k=1 to k=n
            if load[k]= infinite (all servers are all overload)
                put the request into the delay service queue;
            else
                calculate the comparative weight-value of server k and confirm it's
                probability space R(k);
                rand=random();
                if rand belongs to R(k)
                    distribute the request to server k
} //Request_distribution()
Search_Type(){
Switch request type:
    Case issuance type: calculate it's request weight-value; break;
    Case affair type: calculate it's request weight-value; break;
    Case dynamic page type: calculate it's request weight-value; break;
    Case security type: calculate it's request weight-value; break;
    Case multi-media type: calculate it's request weight-value; break;
}

```

3.3. Analysis of QSC-Load Balancing

1. Query of hash table

Since the load equalizer needs to read and write corresponding to records from User_Info_Table, Cluster_Server_Table and Service_Maping_Table, the operation efficiency will effect the system performance. Hashtable is an important storage and index measure and it can query、insert and delete the record efficiently. This algorithm adopts Hash table with bidirectional double linked list to operate the record.

2. Enactment of N_{max} 、T

Enactment of N_{max} 、T is the key problem in this algorithm. The content-based distribution strategy is a exiguous granularity load-balancing one and the better of server performance the larger of N_{max} value. We can set a same N_{max} in the isomorphic server clusters, but in an isomeric system it is necessary to set a different N_{max} according to the server performance. We should consider that an improper N_{max} will make the load unbalanced. The value of T also affects the system performance. The state of each server in the cluster isn't reflected if it is too large, however if it is too small there will be dithering. So the value of N_{max} and T must be generally considered.

3. Dynamic feedback mechanism

The real-time load and response ability of every node are considered in dynamic feedback mechanism. The load equalizer allocates the load according to the real-time feedback information and improves the system efficiency. In this algorithm, the load equalizer checks the load at T intervals and gets the load state via the dynamic feed mechanism.

3.4. Testing of QSC-Load Balancing

Suppose the request is disposed by the method of first in and first out or share time slice. In simulated program, each physical entity is abstracted a Java class, log analysis is an interface class and can be used to dispose different input log. When the load equalizer receives those requests that are disposed by log program, it will distribute them into different servers. The input stream of simulator is a object request one with mark, which can produce asynchronously or via disposing log in server.

Simulate three different request strategies: least_connections strategy, random strategy, QSC-Load Balancing strategy. The throughput contrast of cluster server system is shown as Fig. 2:

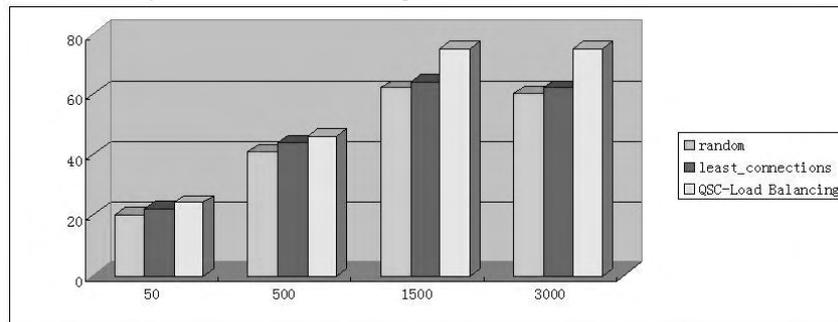


Fig. 1. Throughput contrast of cluster server system

The response time contrast of cluster server system is shown as Fig. 3:

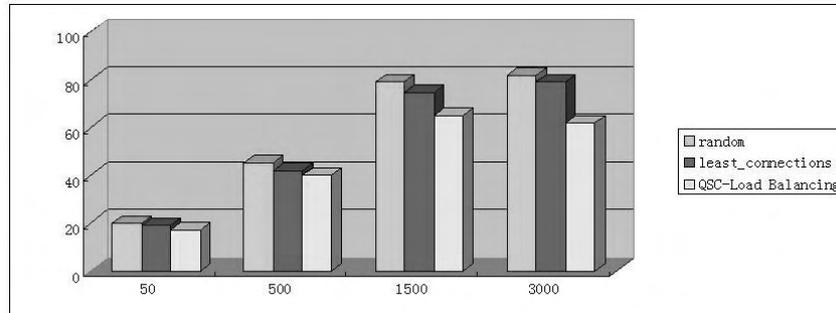


Fig. 3. Response time contrast of cluster server system

Have a research on the load situation of requesting distribution and stimulating server through the analysis of the log on the websites. In the stimulated system every object can be abstracted to a Java class. Log analysis is an interface class, which used to handle different input logs. According to the different request to the load equalizer, this can be equipped within the basic content and maintaining line, will send the request to different server.

4. Conclusions

This paper points out a developed strategy on the basis of the context distribution contra pose to the different requesting content. Show the load condition of every server in the web class more currently through the evaluation of the web content and load server. Load lays the foundation in order to distribute the request balance; introduce the concept of probability-space within the process of distribution to make the load more balance; to eliminate the period related effect due to the adoption of the fixed proportion model. Thus enhance its performance.

5. References

1. Chi-Chung Hui, Samuel T. Chanson. Improved Strategies for Dynamic Load Balancing. IEEE Concurrency, 1999.
2. Emiliano Casalicchio, Michele Colajanni, Scalable Web clusters with Static and Dynamic Contents, University of Rome Tor Vergata, ecalicchio@ing.uniroma2.it.
3. C.-S. Yang, M.-Y. Luo. A content placement and management system for distributed Web-server systems. Proc. of IEEE 20th Int. Conf. on Distributed Computing Systems (ICDCS'2000), Taipei, Taiwan, Apr. 2000.
4. Vivek S. Pai, M. Aron, G. Banga, M. Svendsen, P. Druschel, W. Zwaenepoel, E. Nahum. Locality-aware request distribution in cluster-based network servers. In

Zhang Lin, Li Xiao-ping, and Su Yuan

- Proc. of 8th ACM Conference on Architectural Support for Programming Languages and Operating Systems, San Jose, CA, Oct. 1998.
5. M Arlitt, C Williamson. Web server workload characterization: The search for invariants (extended version). In: ACM SIGMETRICS'96. Marriott, PA, 1996.
 6. Haakon Brhni. A Comparison of Load Balancing Techniques for Scalable Web Servers. University of Oslo EsPen Klovning and ivind Kure, Telenor Research and Development. IEEE Network July/August 2000.
 7. Michele Colajanni, Philip S. Analysis of Task Assignment Policies in Scalable Distributed Web-server Systems, IEEE Transactions on Parallel and Distributed Systems, vol. 9, No. 6, June, 1998.

Zhang Lin, Master, lecturer; research fields: network security, software engineering.

LI Xiao-ping, Ph.D., professor; research fields: network security, information processing, multimedia and image processing.

Su Yuan, Master; research fields: network security.

Received: April 10, 2009; Accepted: September 03, 2009.

...

Blind Separation Using Second Order Statistics for Non-stationary Signals

Jun Du^{1,2}, Ju Liu¹, Shengju Sang^{3,4} and Jun Wang⁴

1. Information Science and Engineering Institute, Shandong University
Jinan, 250100, China
{jundu, juliu}@sdu.edu.cn
2. School of Communication, Shandong Normal University
Jinan, 250014, China
3. School of Mechanical and Power Engineering, East China University of Science and Technology
Shanghai, 200237, China
4. Department of Information Science and Technology, Taishan College
Taian, 271021, China
Sang1108@163.com

Disclosure of Plagiarism

This article is a confirmed plagiarism of the original work: Chia-Chi Lin, "Blind Source Separation For Non-stationary Signals Using Second Order Statistics," Master Thesis, National Tsing Hua University (NTHU), Hsinchu, Taiwan, July 2007.

The Editorial Board of Computer Science and Information Systems sincerely apologizes to the author, Chia-Chi Lin, and his advisors, Professor Wing-Kin Ma¹ and Professor Chong-Yung Chi², for allowing the publication of this article. The article has been removed from the ComSIS Web site, and replaced by instructions to download the original work.

1 Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

2 Institute of Communications Engineering & Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.

Publication information

Volume 7, Issue 1 (February 2010)

Advances in Computer Animation and Digital Entertainment

Year of Publication: 2010

ISSN: 1820-0214

Publisher: ComSIS Consortium

An Accelerometer-Based Gesture Recognition Algorithm and its Application for 3D Interaction

Jianfeng Liu¹, Zhigeng Pan¹, and Xiangcheng Li²

¹ State Key Lab of CAD&CD, Zhejiang University,
310027 Hangzhou, China
{liujianfeng, zgpan}@cad.zju.edu.cn

² Sport System Simulation Lab, China Institute of Sports Science,
100061 Beijing, China
lxc@3s.org.cn

Abstract. This paper proposes an accelerometer-based gesture recognition algorithm. As a pre-process procedure, raw data output by accelerometer should be quantized, and then use discrete Hidden Markov Model to train and recognize them. Based upon this recognition algorithm, we treat gesture as a method of human-computer interaction and use it in 3D interaction subsystem in VR system named VDOM by following steps: establish Gesture-Semantic Map, train standard gestures, finally do recognition. Experimental results show that the system can recognize input gestures quickly with a reliable recognition rate. The users are able to perform most of the typical interaction tasks in virtual environment by this accelerometer-based device.

Keywords: Tri-axes Accelerometer, HMM, Gesture Recognition, 3-D Interaction, Virtual Reality.

1. Introduction

The VR (virtual reality) system has the widespread application in domains as CAD, E-learning, sports simulation, digital entertainment. There are instances of VR system such as VNM [1] (Virtual Network Marathon) and VBL (Virtual Biological Laboratory) developed by State Key Lab. of CAD&CG in Zhejiang University, Virtual Bicycle Training System designed by Sport System Simulation Lab of CISS. The main task of VR system including two aspects: first, organize and manage the virtual scene effectively, and render or describe the virtual objects as clearly as possible; secondly, complete the interaction task between the user and the virtual objects accurately [2].

Benefit from the rapid development of 3D realistic graphics in recent decades, the present VR systems are able to render large-scale virtual scene easily, and give the users strong immersion. Another primary task of VR system is 3D interaction in virtual environment. D. Bowman [2], [3] et al. defined "3D interaction" as: "Human-computer interaction in which the user's

tasks are performed directly in a 3D spatial context.” Different with traditional WIMP style interaction in 2D program, 3D interaction requires a richer interaction techniques and multi-modal user interface to complete various complicated interaction tasks in VR system which always cope with huge media. In order to enrich 3D interaction techniques in VR system, we presented an accelerometer-base interaction technology and applied it in an instance of VR system as a principal interaction device.

As a novel interaction method, gesture recognition has been researched by many researchers. Hofmann, F. et al. [4] presented the velocity profile based method which utilize HMM to recognize human gestures. Portillo-Rodriguez [5] et al. proposed a FSM based recognition method. Jani Mantyjärvi et al. treated the gesture recognition technology as an interactive method in a design environment [6], and developed a system named Smart Design Studio. Based on those previous works, Thomas Schlomer et al. [7] implemented the accelerometer-based gesture recognition algorithm and utilized the low-cost Wii controller as an interaction device in their experiments.

2. Accelerometer based gesture recognition

We utilize a sensor which integrated a tri-axes accelerometer chip as a hand held input device in our interaction system. When the human performs a gesture, the sensor will collect the data flow output by accelerometer chip, and send it to PC via wireless protocol. We consider this raw data stream fetched from sensor as an “input pattern”.

Definition 1. Patten $P = \{V_t \mid 0 \leq t \leq T\}$, where V_t is the acceleration vector output by interaction device at time t , T is the time when data stream terminates.

According to the daily experience, the patterns generated by the movement of hand when human performing the same gestures satisfy certain statistical rules to some extent, based on it we propose the “standard pattern”. The “standard pattern” is a class of pre-defined patterns, each one corresponding to a special “input semantics”. When user performed a gesture, the sensor will send the “input pattern” to interaction system, then system will find out the most approximate “standard pattern”, this also can be regarded as a procedure of recognition, and finally the interaction system get an input semantic according to the recognition result, system will interpret (execute) it and return a feedback information to the user.

2.1. Establish patterns and Preprocess

The accelerometer chip BOSCH SMB380 is selected in our system which has a high sampling frequency and sensitivity. The raw data output by the chip is

noisy, redundant, and approximately continuous. It is too complex to process them directly for our system. Here, a denoising procedure is applied.

When denoising, all elements V_t in input pattern P will be processed by a smoothing function like that:

$$S(V_t) = \begin{cases} V_t & t = 0 \\ S(V_{t-1}) + \alpha(V_t - S(V_{t-1})) & 0 < t \leq T \end{cases} \quad (1)$$

α is a smoothing factor which in the range from 0 to 1.

A quantification process is necessary, because we use the discrete HMM-based approach. It requires that the quantified results should simply and countable, in other words it requires a codebook its size is fixed empirically. Since the quantification targets are 3 dimension vectors, and the correlation between them can be measured by the Euclidean distance. Partitional clustering method is suitable. We choose k-medoids clustering rather than k-meaning clustering, because it is more robust to noise and outliers as compared to k-means [8].

Consider that gestures in our system are performed in 3D space, and the output data is 3D vectors which imply the direction and velocity of the hand's motion, the cluster centers should be in 3D space, rather than 2D circle in most gesture recognition based on vision method [9]. Thomas Schlomer et al. [7] distribute the cluster centers in 3d sphere. In this paper we rely on this idea, and identified k (the number of cluster centers) = 14. Then perform clustering algorithm [8]:

Before clustering, for every input pattern P , calculate the maximum and minimum normal of its element vectors. Thus, the radius of the initial clustering sphere can be identified as $(\max \text{ normal} + \min \text{ normal})/2$. It will make clustering quickly.

1. First, distribute k initial medoids uniformly in a 3D sphere. The initial medoids set $M_{ini} = \{m_1, m_2, \dots, m_k\}$ can be identified according to the $R_{cluster}$.
2. Associate each vector in the input pattern to most similar medoid. The similarity here is defined using distance measure that is Euclidean distance.
3. Randomly select nonmedoid object O' .
4. Compute total cost S of swapping initial medoid object to O' .
5. If $S < 0$, then swap initial medoid with the new one (if $S < 0$ then there will be new set of medoids).
6. Repeat steps 2 to 5 until there is no change in the medoid.

After clustering, each candidate vector is associated to an indexed cluster. Quantification is on the basis of clustering results, the size of the codebook obviously equals to the number of clusters. For a vector (component in a pattern), identify its quantification result as the cluster index which it belongs to.

In summary, for a given "input pattern" P , after the steps of denoising and clustering, a desired pattern $P' = \{ID_{V_i} | V_i \in P\}$ which is suitable for following steps can be established.

2.2. Patten classification

Statistical pattern recognition is based on statistical characterizations of patterns, assuming that the patterns are generated by a probabilistic system. A wide range of algorithms can be applied for pattern recognition, such as Support Vector Machine, Neural Network, and Hidden Markov Model. Since our pattern is composed of a set of discrete and simply data, we select one dimension and discrete HMM in our approach.

A HMM can be formulized as $\lambda = \{N, M, \pi, A, B\}$ [10]. N is the state set, M is the observation value set, π is the probability vector of initial states, A is the matrix for the state transition probability distributions, and B is the matrix for the observation symbol probability distributions.

Training

Definition 2. Gesture G is denoted as a two-tuple $G = \langle P, Att \rangle$, where P is input pattern when perform a gesture, and Att is a set of additional information, such as input data provided by any other device at the same time.

In this paper, the item “standard gesture” means a class of gestures which correspond to a special “standard pattern”. Every “standard gesture” is mapped to an interaction semantic. When system receives an input gesture, the system should give a “standard gesture” most approximate to input gesture.

For each “standard gesture”, our participants perform it repeatedly. So finally we get a set of training data $T = \langle G_1, G_2, \dots, G_n \rangle$ for every “standard gesture”.

Training approach as follows:

1. Establish a HMM for a “standard gesture” and initialize it.
2. Compose the multi-dimensional training data set $\{O_1, O_2, \dots, O_L\}$, where O_i is the first component P in G_i .
3. Consider that each elements in T are independent events, $P(O|\lambda)$ can be expressed like that $P(O|\lambda) = \prod_{i=1}^L P(O_i|\lambda)$ where L is the length of T . Reestimate the factors $\{A, B, \pi\}$ of HMM by Baum-Welch algorithm [10], and get a new HMM denoted as λ' .
4. If $P(O|\lambda') - P(O|\lambda) > \varepsilon$, $\lambda = \lambda'$, then repeat step 3. If $P(O|\lambda') - P(O|\lambda) \leq \varepsilon$, λ' can be regard as the local optimal solution.

Finally, a set of optimized HMMs correspond to “standard gesture” can be obtained.

Recognition

The Recognition process might regard as gains an arbitrary input gesture (pattern), and find the optimal matching gesture (pattern) in the “standard gesture” database.

Formal description of the recognition process is as follows:

1. For a given input gesture $G_x = \langle P, Att \rangle$, P was treated as an observation value sequence O .
2. Evaluate $\sigma_n = \arg \max P(O | \lambda_i)$, λ_i is one of the trained HMMs corresponding to “stand gestures”.
3. Thus, the λ_i that give the maximum evaluation of σ_n can be considered as the result. And the input gesture was recognized.

The recognition system outputs a “standard gesture” in the database which the result HMM corresponding to.

Evaluation with a threshold

Think about that, for any input gestures, the above algorithm will give a solution. But it cannot guarantee that it is a optimal and harmless solution, since when user performed a misoperation and the system will recognize it and do some work that violate the user’s original intention. To avoid it, a threshold should be set when evaluating the maximum probabilities. It can be regarded as a misoperation and ignore the input when the maximum expectation we have got in “recognition” step is not greater than the threshold empirical identified.

3. 3D Interaction System in Virtual Environment

3D human-computer interaction in virtual environment typically includes three main tasks [11]: virtual objects manipulation, viewpoint manipulation, application control. Viewpoint manipulation including the movement of viewpoint (et. Camera) and the control of other parameters such as FOV and Zoom [12]. The term application control means the communication between user and system which is not part of virtual environment.

We propose a 3D interaction model in virtual environment. Its goal is to complete above tasks effectively.

Before that, some concepts should be introduced:

Definition 3. Interaction Atom was denoted as A , means an atomic interaction operation. A^* stands for the Interaction Atom set which should be determined early in the design of interaction system.

Definition 4. Execution Routine $E = \{A_0, A_1, \dots, A_t\}$, A_i is an interaction atom, and the entire routine was composed by a sequence of atomic interaction operations to implement a special interaction task.

Definition 5. Interaction Semantic is a four tuples as $s = \langle I, C, E, con \rangle$, where I is the input data flow, C is the current context and E is the target execution routine and will be executed only if the condition con is true.

The architecture of the model is illustrated in Fig.1.

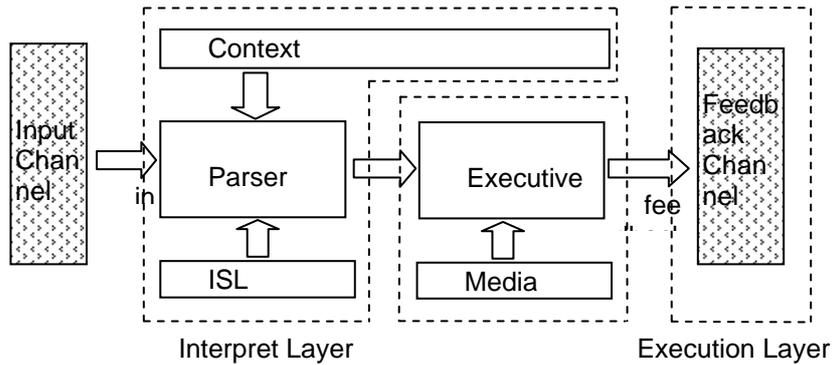


Fig. 1. The architecture of interaction model.

3.1. Interaction Semantics Library

The term “Interaction Semantic” refers to the “meaning” of the input data flow in particular context. The phrase “meaning” can be considered as a sequence of interaction commands and their parameters. Once an input event occurs, the Parser will interpret the input data flow, and extract the underlying Interaction Semantics in it by referring to current global context. The Executive receives Interaction Semantics from Parser, translates the semantics to Execution Routines, and completes the interaction tasks by executing the routines.

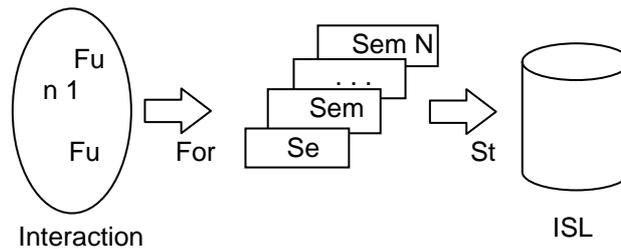


Fig. 2. creating interaction semantic library.

An ISL (Interaction Semantic Library) must be established when designing interaction system. Enumerates all interactive functions, translates them into IS, and eventually stored the ISs into ISL. The process was illustrated in Fig.2.

3.2. Interaction Feedback

The interaction result feedback is an important part of interaction process. A good interactive system needs clear and complete feedback. The feedback message must contain result and final state of the task, the change of global context caused by the last interaction, the reasons when task failure, as well as some proposal of alternative plan.

4. Applications and Analysis

We applied 3D interaction method based on the gesture recognition algorithm which introduced above in the VR system –Virtual Digital Olympic Museum (VDOM).The 3D Interaction subsystem was demonstrated in Fig.3.

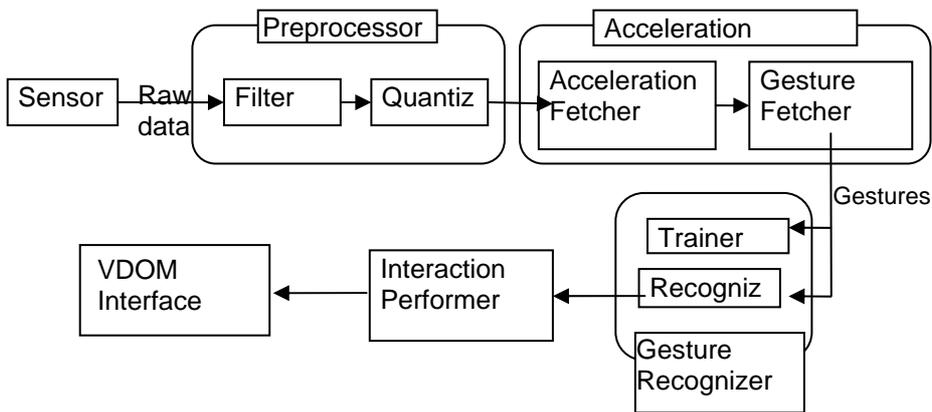


Fig. 3. 3D-Interaction Subsystem in VDOM.

4.1. Gesture-Semantic Mapping

In VDOM, interaction system provides interaction functions as follows:

<1>Agent control: turn left/right

<2>Object manipulation: selection/rotation/scaling/movement

For the functions, ISL was established. In the system, ISL was saved as XML file and named *.isl.xml. It is facile to deal with by script language.

Table.1 gives the mapping from “standard gesture” to “interaction semantic” where gestures (A)-(H) was illustrated in Fig.4.

Table 1. Gesture-Semantic Table

Gesture	Interaction Semantic
A	move left/turn left
B	move right/turn right
C	rotate left
D	rotate right
E	move up
F	move down
G	confirm/select/zoom in
H	cancel/delete/zoom out

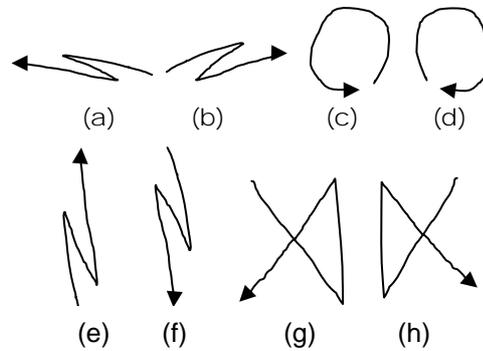


Fig. 4. Gestures.

4.2. Training

Table.2 shows the counter of training samples for each gesture, for gesture g (confirm) and h (cancel) are more complex than other gestures in Fig.3, more samples is helpful. After trained by the samples, we gain 8 HMMs corresponding to gestures A-H.

The training performance can be measured by $P(G | \lambda)$ which stands for the probability of HMM λ produce a observation sequence G . Fig.4 shows $\ln P(G | \lambda)$ with a varying number of HMM states. Training performance of each gesture was enhanced while increase the HMM State count from 8 to 12. While increase HMM count from 12 to 16, we gain an improvement in performance for gesture a,b and d, however we also got a loss in performance for gesture c,f,e. The count was identified as 12 by considering with both training effect and speed.

Table 2. Training Data Table

Interaction Commands	Training Data
move / turn left	150
move / turn right	150
rotate left	150
rotate right	150
move up	150
move down	150
Confirm	200
Cancel	200

Table 3. Recognition Results

Interaction Commands	Testing Data	Correct	Recognition Ratio
move / turn left	50	47	0.94
move / turn right	50	46	0.92
rotate left	50	49	0.98
rotate right	50	47	0.94
move up	50	48	0.96
move down	50	49	0.99
confirm	50	44	0.88
cancel	50	45	0.90

4.3. Results

After completing the training work, test each gesture with 50 samples separately. And the result was shown in Table.3. For gesture g and h, we got a recognition rate slightly less than 90 percents on average, and for the remaining we got a rate more than 95 percents.

As a whole, the 3D interaction technology based-on accelerometer was applied in VDOM successfully, and the system shown a reliable result. Users can use the accelerometer-based device barrier-free and perform the interaction with VDOM easily. Fig.6 contains the screenshots of VDOM. In the left side, the user is performing gestures which defined in Fig.4 with our accelerometer-based device, and the interaction results are illustrated in the right side.

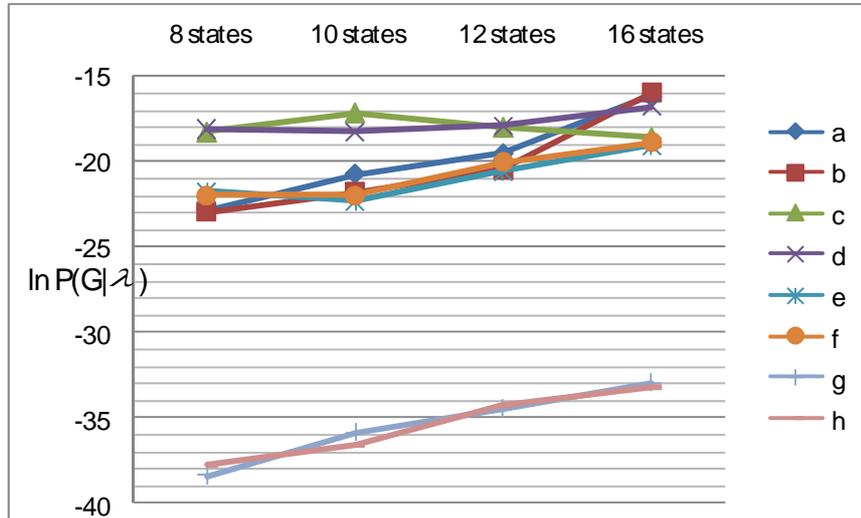


Fig. 5. $\ln P(G|\lambda)$ with a varying number of HMM states.

5. Conclusion

We have presented an accelerometer-based human-computer interaction method and its application in VR system. We also described the implementation of interaction subsystem. A Gesture-Semantic Table contains all predefined standard gestures and their corresponding semantics. Semantics and their respective execute routines are stored in ISL. Based on Gesture-Semantic Table and ISL the interaction subsystem can identify standard gestures in input data stream, and translate them into interaction semantics and perform the interaction tasks finally. In our interaction subsystem, there are several reasons may cause a lower recognition rate: first, a deficient training process causes weakly correlation between HMM and the "standard gestures"; secondly, for complicated gestures, it may lose parts of information during quantification process; thirdly, the lack of samples from various users. In practical application, these following ways can improve the recognition rate: increase the sample count, involving various participants; increase the size of the codebook appropriately, in other words, increase the count of cluster centers in the clustering step, it may reduce the loss during quantification; for complicated gestures, increase the number of HMM state. The future work is to extend current gesture library and ISL, and to provide more interaction functions in 3D virtual world.

An Accelerometer-Based Gesture Recognition Algorithm and its Application for 3D Interaction

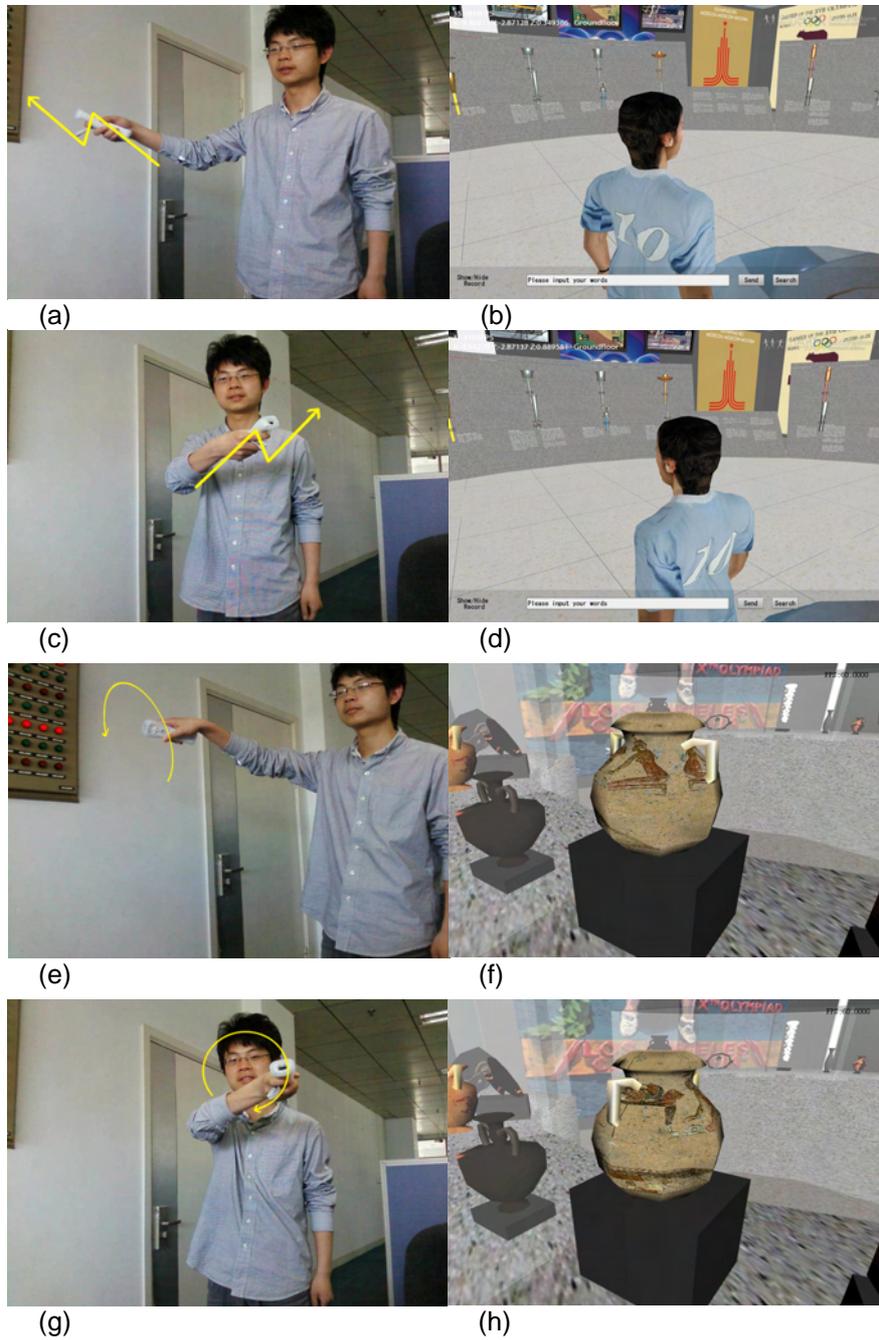


Fig. 6. Runtime Results.

6. References

1. He GQ, Pan ZG, Li YQ, Su SY: Synthesized Synchronicity Mechanism for Fitness-Oriented Virtual Network Game. *Journal of Computer-Aided Design & Computer Graphics*, Vol.12, No.2, 73-78. (2008)
2. Bowman, D.A., Kruijff, E., LaViola, J., Poupyrev, I.: *3D User Interfaces: Theory and Practice*. Addison-Wesley, Boston, USA. (2004)
3. Bowman, D.A., Coquillart, S., Froehlich, B.: New Directions in 3D User Interfaces. *Computer Graphics and Applications, IEEE*, Vol.28, No.6, 20-36. (2008)
4. Hofmann, F., Heyer, P., Hommel, G.: Velocity Profile Based Recognition of Dynamic Gestures with Discrete Hidden Markov Models. *International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction*, Springer, London, UK, 81–95. (2004)
5. Portillo-Rodriguez, Sandoval-Gonzalez, O.: Development of a 3D real time gesture recognition methodology for virtual environment control. *The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 279-284. (2008)
6. Jani Mantyjarvi, Juha Kela, Panu Korpipaa.: Accelerometer-based gesture control for a design environment. *Personal and Ubiquitous Computing*, Vol.10, No.5, 285-299. (2006)
7. Thomas Schlomer, Benjamin Poppinga, Niels Henze, Susanne Boll.: Gesture recognition with a Wii controller. *Proceedings of the 2nd international conference on Tangible and embedded interaction*, 11-14. (2008)
8. Ng, R.T., Han, J.: Efficient and effective clustering methods for spatial data mining. *Proceedings of the 20th VLDB Conference*, 144-155. (1994)
9. Hyeon-Kyu Lee, Jin H.Kim.: An HMM-based threshold model approach for gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.21, No.10, 961-973. (1999)
10. Rabiner, L.R.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE*, Vol. 77, 257-285. (1989)
11. Chris, H.: A Survey of 3D Interaction Techniques. *Computer Graphic Forum*, Vol.16, 269-281. (1997)
12. LaViola, J., Acevedo, D.: Hands-Free Multi-Scale Navigation in Virtual Environments. *Proceedings of ACM Symposium on Interactive 3D Graphics*, North Carolina, USA, 9-15. (2001)

Jianfeng Liu was born in 1985. He is a M.S. candidate at the College of Computer Science and Technology, Zhejiang University. His current research interests include virtual reality and human-computer interaction.

Zhigeng Pan was born in 1965. He is a professor and doctoral supervisor at the College of Computer Science and Technology, Zhejiang University. His researches areas are virtual reality, argument reality.

Xiangcheng Li was born in 1974. He is currently the Director of Sport System Simulation Lab of CISS. His researches areas are virtual reality, sport simulation.

Received: April 28, 2009; Accepted: July 15, 2009.

The Design and Evaluation of Hierarchical Multi-level Parallelisms for H.264 Encoder on Multi-core Architecture

Haitao Wei¹, Junqing Yu¹, and Jiang Li¹

¹School of Computer Science & Technology,
Huazhong University of Science & Technology, 430074 Wuhan, China
yjqing@hust.edu.cn

Abstract. As a video coding standard, H.264 achieves high compress rate while keeping good fidelity. But it requires more intensive computation than before to get such high coding performance. A Hierarchical Multi-level Parallelisms (HMLP) framework for H.264 encoder is proposed which integrates four level parallelisms – frame-level, slice-level, macroblock-level and data-level into one implementation. Each level parallelism is designed in a hierarchical parallel framework and mapped onto the multi-cores and SIMD units on multi-core architecture. According to the analysis of coding performance on each level parallelism, we propose a method to combine different parallel levels to attain a good compromise between high speedup and low bit-rate. The experimental results show that for CIF format video, our method achieves the speedup of 33.57x-42.3x with 1.04x-1.08x bit-rate increasing on 8-core Intel Xeon processor with SIMD Technology.

Keywords: H.264 encoder; Hierarchical Multi-level Parallelisms; Multi-core Architecture.

1. Introduction

H.264 [1] as a video coding standard is now being used widely due to its high-quality video content and low bit-rate. However, it makes encoding process more complex and requires more computation than previous coding standards. Given fixed fidelity, H.264 reduces bit-rate up to about 50% at the cost of more than three times computational complexity compared to H.263 [2]. Therefore, the hardware and software co-design parallelisms are needed to accelerate the speed of encoder for real-time application. Multi-core processor architecture [3] is now becoming the mainstream solution for next generation general computation. Unlike the simultaneous multiple threading (SMT) [12] and hyper-threaded processor (HT) [13] where most micro-architectures are shared between logical processors, multi-core processor introduces new microprocessor technologies to deliver high computation ability. First, multi-core processor integrates multiple single processor cores into one chip which

supports the real coarse-grained hardware thread parallelism. Second, each core is equipped the SIMD instruction sets to provide the fine-grained parallelism. Third, each core has independent L1/L2 cache to increase the bandwidth and hit rate. All these features can be beneficial for improving the speed of H.264 encoder.

Many parallel algorithms for H.264 encoder were discussed in previous work. A parallel scheme is addressed in [4, 5] that encodes the slices of a frame in parallel on Intel hyper-threading architecture. It mainly concentrates on the slice parallelism based on fixed IBBP encoding structure. A method that utilizes the dependency of reconstructed macroblock (MB) and encoding macroblock to encode multiple macroblocks in parallel is reported in [6]. Another parallel algorithm for macroblock encoding is reported in [2, 11]. It uses approximate neighboring encoding information to find the optimal coding mode of the current coding block. A pipeline algorithm is discussed to parallelize macroblock analysis and the performance is analyzed on Cell processor in [7]. A H.264 decoder is implemented on general-purpose processors by using SIMD instructions in [8]. Parallel motion estimation scheme for H.264 are discussed in [9, 10].

We expand the method proposed in [4, 5] to multi-B frames in the frame level and combine frame, slice, macroblock and data parallelisms for H.264 encoder into one HMLP framework. First, the HMLP model and the design details of each level parallelism for H.264 encoder are presented. Then, based on performance analysis on each parallelism the tradeoffs between multiple parallel levels are attained to optimize the encoding performance.

The rest of this paper is organized as following. Section 2 provides detail design and implementation of our HMLP model for H.264 encoder. Section 3 demonstrates performance results and discusses the results. The selection strategy of multi-level parallelisms is illustrated in section 4. And section 5 concludes this paper.

2. Hierarchical Multi-level Parallel Parallelisms for H.264 Encoder

In H.264, a video sequence includes many frames. Each frame is partitioned into slices, which is the encoding unit and independent of other slices in the same frame. Slice can be decomposed into macroblock which is the unit of encoding algorithm. The structure above provides potential parallel optimization opportunities.

2.1. The Framework of HMLP model

As in Figure 1, the framework of HMLP model for H.264 encoder is designed to integrate four levels of parallelisms of frame, slice, MB and data into one implementation. It consists of encoding threads and queue buffers. Three

kinds of encoding thread – frame thread, slice thread and MB thread, do the encoding process at three different levels. Frame thread is on the top level. Frame threads create the threads for the slice-level which hierarchically create the threads for the MB-level. The data-level parallelism which acts as functional parallelism is included in the MB encoding thread. All above parallelisms compose a hierarchical parallelism tree, where from root node to the leaf node the parallelism grain is decreasing. The HMLP framework shows good scalability. In each parallel level, the size of the processing unit for each thread can be decreased to increase the number of thread. For example the frame can be decomposed into more slices to increase the slice encoding thread. For different levels, because of the hierarchical structure of frames, slice, MBs and data there are many parallel grains to select the size of processing unit.

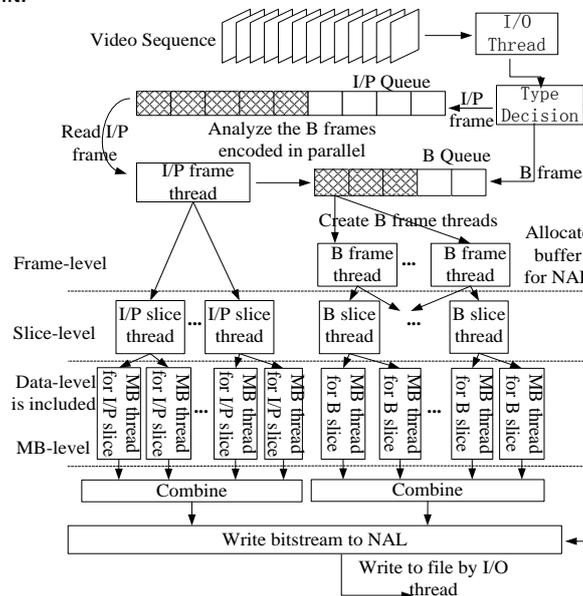


Fig. 1.. Hierarchical Multi-level Parallelisms (HMLP) framework for H.264 encoder.

2.2. Frame and Slice level Parallelisms Design and Implementation

Usually, a sequence of frames is encoded using an IB...BPB...BP... structure. The number of B frame between two P frames can be multiple. Here, I and P frames are treated as the reference frames and B frame are considered as non references in order to explore more parallelism. Figure 2 shows the principle of frame-level parallelism. The display order indicates the original order of video frame. The dependency between the frames is showed in the encoding order. In this encoding order, the completion of encoding a P frame will make the subsequent one P frame and some B frames ready for encoding

in parallel. Here, one P frame and the B frames in the same column will be encoded in parallel order.

H.264 encoder is divided into three parts: input processing, encoding and output processing. As depicted in figure 1, the input processing reads uncompressed images, decides type and allocates the NAL node for bit-stream. The output processing checks the NAL queue and writes the bit-stream after encoding to the output file. One I/O thread is used to handle the input and output processing. In order to explore parallelism in between frames, two queues are used, one is for I or P frames and another is for B frames. After each frame's type is decided the frame will be put into the corresponding queue. The I/P encoding thread will fetch an I or P frame from the I/P queue and check the B frames which are independent of current I/P frame and ready for encoding in the B queue. After that, I/P thread will create B frame encoding thread for each above B frame and encode the I/P frame with these B frames in parallel.

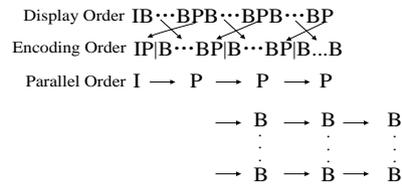


Fig. 2. The frame dependencies and parallel encoding order in H.264.

A frame can be divided into small slices which are independent and can be encoded in parallel. As figure 1 illustrates, each frame encoding thread like I/P thread and B thread divides the frame into slices and create encoding thread for each slice. After encoding, each slice thread writes the bit-stream to the NAL in the order of slice.

The pseudo code of frame-level and slice parallelisms is listed below. We use I/O thread to process the input and output and I/P thread to create B frame encoding threads dynamically to encode the frames in parallel. In each frame encoding, frame is partitioned into slices and slice encoding threads are created for each of it to encode. Finally, the bitstream is assembled and write to the file.

I/O thread Code:

```

while (input video sequence is not NULL) do
  if (there is free entry in image buffer) then
    read a frame to image buffer and decide its type;
    if (the type is I or P) then
      enter I/P queue;
    else
      enter B queue;
    end
    allocate a node in NAL queue for current frame;
  else if (there is bitstream node in the NAL queue)
    write the bitstream node to output file;
  else wait;

```

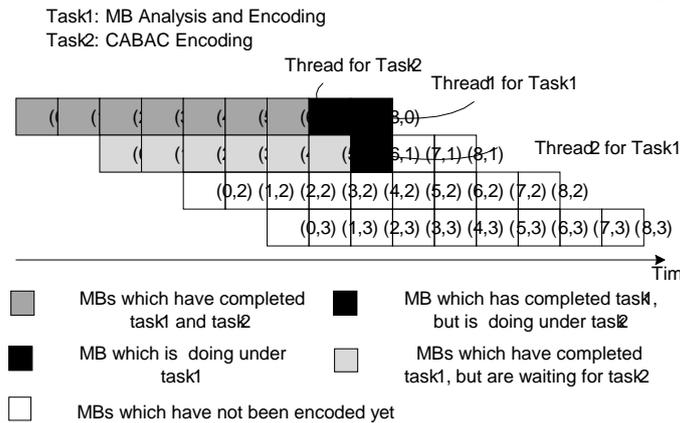
```
end
end while
I/P thread Code:
while (true) do
  if (there is frame in the I/P queue) then
    fetch a frame from I/P queue;
    analyze the B frames in B queue;
    create Encoding thread for B frame
      which can be encoded in parallel;
    call Encoding thread to encode current frame;
  else if (all frames are encoded) exit;
  else wait;
  end
end while
Encoding thread Code:
for each slice in the frame
  create slice encoding thread to encode;
assemble the bitstream and write to file
```

2.3. MB and Data Level Parallelisms Design and Implementation

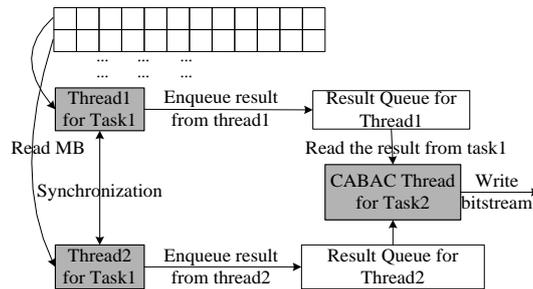
MB encoding process is the most time-cost part in H.264. In the implementation, it is composed of three modules – MB analysis, encoding and CABAC. MB analysis module mainly analyzes intra and inter prediction mode, predicts motion vector, and decides the MB type. And the MB encoding module mainly processes the DCT, quantization and de-blocking filtering. From the analysis in [6], the processing above indicates that the MB analysis and encoding of current MB depends on results of current MB's top and left neighboring MBs. The CABAC module depends on the CABAC result of last MB. So, it must be processed sequentially in the row order.

Figure 3(a) illustrates the principle of parallel MB encoding process in a slice structure that consists of 4x9 MBs. MB is indexed by the coordinate MB(i,j). According to the dependency, in order to analyze and encode MB(i,j), it is necessary to refer to the results of its left MB(i-1,j), top-left MB(i-1,j-1), and top MB(i,j-1). Therefore, the initial way is to analyze and encode two MBs, MB(0,0) and MB(1,0) sequentially. After this, MBs in each column such as MB(2,0) and MB(0,1) can be analyzed and encoded in parallel, meanwhile, MB(0,0) and MB(1,0) can do CABAC. And then MB(3,0) and MB(1,1) can be analyzed and encoded in parallel. In such a way, MB(6,0), MB(4,1), MB(2,2) and MB(0,3) can be analyzed and encoded in parallel. However, CABAC must be processed sequentially in the row order. According to the Amdahl's law, the total time is decided by the cost time of CABAC. The experiment shows the cost time of CABAC is about half of the sum of MB analysis and encoding time. So, as figure 3(a) illustrates, three threads are created for MB encoding process, two of which execute MB analysis and encoding (Task1) and one of which executes CABAC (Task2). The producing rate of two threads for Task1

is enough to match the consuming rate of one thread for Task2. In the parallelization pattern, each slice is partitioned into MB rows. Each thread for Task1 processes the interlacing MB rows in a slice – one thread processes odd rows the other processes even rows. Thread for Task2 processes MB rows in sequential and synchronize with the two threads executing Task1.



(a) Principle and task partition for parallel MB encoding process



(b) The implementation of MB-level parallelism

Fig. 3. The multi-threads implementation of macroblock-level parallelism

Figure 3(b) shows the implementation of MB-level parallelism. The slice thread created in section 2.2 is taken as the CABAC thread which creates two threads for MB analysis and encoding. Two queue buffers are used to store the results of MB analysis and encoding from the threads. CABAC thread reads the two queue buffers alternatively to encode and write the final results to bit-stream.

The SIMD technique can be used to speed up encoding process in the data-level. We use the SIMD instruction to rewrite the following encoding modules: integer DCT/IDCT transform, quantization, motion compensation, sub-pel search, de-blocking and SAD calculation. Because the SIMD is an instruction optimization technology, it does not compete with frame or slice parallelism for physical threads.

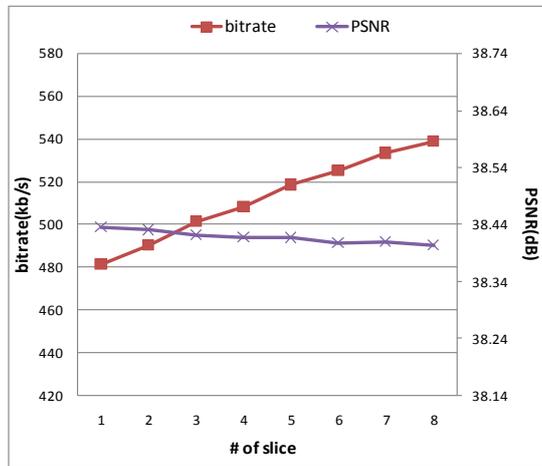
3. Experiments and Performance Evaluation

The experimental tests of multi-level parallel H.264 encoder is performed on 8 cores Intel Xeon processor running at 2.0GHz, 1M L2 Cache and supporting MMX/SSE1/SSE2. If it is unspecified, the test video is Foreman in CIF format (352x288) with 300 frames. The profile of H.264 encoder is main profile which is configured as following: (1) inter-coding using B-slices and weighted prediction; (2) deciding references on a per partition basis; (3) using hexagonal search; (4) using 1/4-pel resolution search (5) enabling all search types; (6) using CABAC.

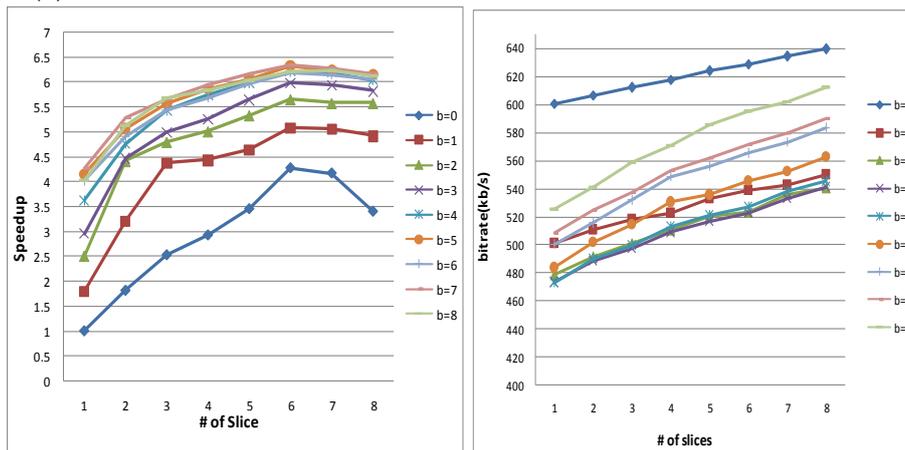
3.1. Coding Performance Versus Frame and Slice Parallelisms

The coding performance is one of the most important issues in the video coding. Even though parallelisms can make video data process faster, it must not significantly sacrifice the coding performance. Figure 4(a) shows the encoder performance when a frame is divided into different number of slices, here number of B-frame is 2. We can see that with 8 slices in each frame, we have a bit-rate increment close to 15% which is not admissible. Too much slice parallelism causes bit-rate rising. Thus, the slice parallelism is sensitive and restricted to bit-rate. Another quality parameter PSNR does not behave so adversely. It is seen that PSNR has a small variation around 38.42 dB. It is concluded that *bit-rate is the key coding performance parameter* that limits frame and slice parallelisms.

As mentioned early, B frame can be encoded with P frame in parallel, so multiple B frames can increase the degree of parallelism. But, it also influences the bit-rate and drops down the image quality because of the inaccurate bi-predictions. One challenge is to attain a high quality. So, the proper amount of B frame should be selected. Meanwhile, partitioning one frame into multiple slices can increase the degree of parallelism, but it also increases the bit-rate. Because it isolates the correlation between different slices in one frame and adds slice heads to the bit-stream. Thus, the amount of slice should be selected carefully as well. Figure 4 illustrates the speedup and bit-rate variation with of the number of B frames and slices. In figure 4 (b), There is a best speed up of 6x to 6.3x when the number of B frames ranges from 3 to 7 and the number of slices in each frame is 6. In figure 4 (c), the bit-rate descends about 110kb/s when the number of B frames ranges from 0 to 3 and rises up about 50 kb/s when B frames varies from 3 to 8. Thus, considering frame level only, best speedup and relative lower bit-rate are achieved when the number of B frames ranges from 2 to 3. On the other side, given 2 or 3 B frames, the bit-rate increases almost linearly with the number of slices. The bit-rate increases about 40kb/s compared with no slices partition when the number of slices reaches 6, at which the best speedup is attained. *The important observation is that setting the number of B frames to 2 to 3 and partitioning a frame to 6 slices delivers the best tradeoff for frame and slice parallelisms that achieves a 6.0x-6.3x speedup with 1.08x bit-rate.*



(a) bit-rate and PSNR vs. the number of slices in a frame



(b) speedup

(c) bit-rate

Fig. 4. Speedup and coding performance of frame-level and slice-level parallelisms.

3.2. Coding Performance Versus MB and Data Parallelisms

As mentioned before, MB-level parallelism utilizes the inherent dependencies of different MB encoding processes in a slice. Thus, it can increase the degree of parallelism while keeping the bit-rate no changing. Figure 5 shows the speedup of adding the MB-level parallelism to frame-level and slice-level parallelisms. Comparing to the number of 6 slices where the best speedup is achieved in figure 4 (b), the best speedup in figure 5 shifts to point of 3 slices. Meanwhile, the speedup of frame-level parallelism almost keeps the same.

MB-level parallelism doesn't increase the bit-rate. As refers to figure 4 (c), the bit-rate decreases about 20 kb/s when the number of slices reduces from 6 to 3 where the peak speedup is achieved. *One important conclusion is that MB-level parallelism decreases the bit-rate while keeps the best speedup of 6.x through reducing the number of partitioned slices and increasing the MB parallelism in a frame.* It is observed that when number of B frame is 2 to 3, the partitioned slices in a frame is 3 and the MB-level parallelism is used we can achieve a good speedup and maintain a lower bit-rate.

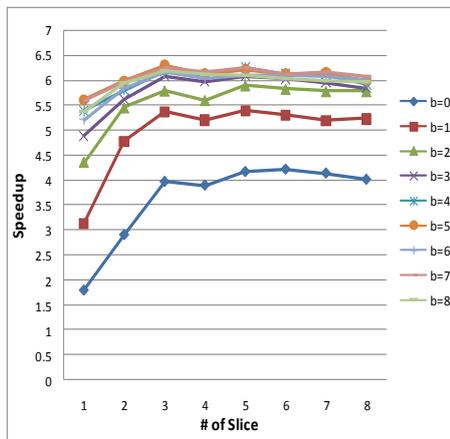


Fig. 1. Speedup of adding MB-level parallelism to frame-level and slice-level

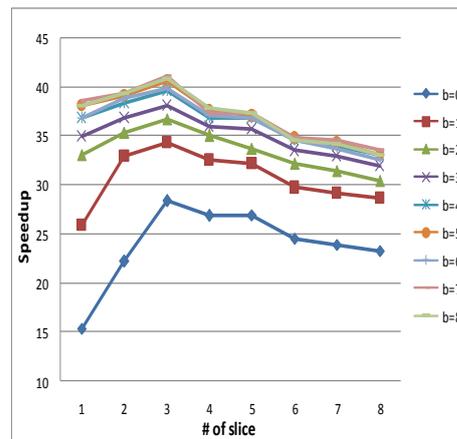


Fig. 2. Speedup of combining four level parallelisms of frame, slice, MB and data.

Data-level parallelism utilizes the SIMD instruction to improve the computation of encoding process especially the vector and matrix computation. Thus, it will not increase the bit-rate as well. As figure 6 illustrates, about 42.3x speedup is achieved with 1.04x bit-rate arising by combining four level parallelisms.

3.3. Performance Comparison with Other Related Works

Table 1 shows the performance comparison between our hierarchical multi-level parallelism H.264 encoder with other related works. In [5], slice-level parallelism is used to a fixed frame structure. For 2 B frames the speedup of 4.31x-4.69x is achieved on 4 Intel Xeon processors with Hyper-Threading Technology (8 logical processors). This method is implemented in our test bed and achieves the speedup of 5.56x-5.72x while with 1.11x bit-rate (ratio). Single macroblock-level parallelism method in [6] and single data-level parallelism method in [8] achieve the speedup of 3.08x and 2x-4x separately, and keep the bit-rate no change. Comparing with above method, for 2 B frames structure, our hierarchical multi-level parallelisms method gains the

speedup of 33.57x-34.78x while with 1.05x bit-rate (ratio). Our multi-level method replaces part slice-level parallelism with the macroblock-level parallelism to reduce the number of slices.

Table 1. Performance comparison with other works

Parallelization method	Speedup	Bit-rate(ratio)
Slice-level with fixed B frame[10,18]	5.66x-5.72x	1.11x
Single MB-level parallelism[5]	2x-4x	1x
Single data-level parallelism[17]	3.08x	1x
Hierarchical multi-level parallelisms in our work	33.57x-34.78x	1.05x

4. Conclusions

H.264 provides many potential parallel optimization opportunities. Single level parallelism scheme can speed encoding, however, it achieves low speedup and increases the bit-rate. A hierarchical multi-level parallelisms design for H.264 encoder is presented which exploits the multi-level parallelisms of frame, slice, macroblock and data in one implementation on multi-core architecture. The tradeoffs of integrating multiple levels are analyzed to gain good speedup and also to keep bit-rate and the video degradation as minimal as possible. The speedup of 42.3x is achieved on 8 Intel SIMD processors with SIMD Technology The method demonstrated can also be applied to other video coding and parallel hardware.

5. Acknowledgement

This paper is financially supported by the National Natural Science Foundation of China under Grant No.60703049; Intel grant for a study of multi-core programming environment.

6. References

1. T. Wiegand, G. J. Sullivan, G. Bjontegaard et al.: Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No. 7, 560-576. (2003)
2. T-C. Chen, Y-W. Huang, and L-G. Chen.: Analysis and design of macroblock pipelining for H.264/AVC VLSI architecture. In *Proceedings of 2004 International Symposium on Circuits and Systems*, Vol. 2, 273-276. (2004)

3. Jeff Parkhurst, John Darringer, PBill Grundmann.: From Single Core to Multi-Core: Preparing for a new exponential. In Proceedings of the 2006 IEEE/ACM International Conference on Computer-aided Design, 67-72. (2006)
4. Chen Yen-Kuang, Tian Xinmin, Ge Steven and Girkar Milind.: Towards Efficient multi-level threading of H.264 encoder on Intel Hyper-Threading Architectures. In Proceeding of International Parallel and Distributed Processing Symposium. Vol.18: 889-898, (2004)
5. Ge S., Tian X., and Chen Y.-K.: Efficient Multithreading Implementation of H.264 Encoder on Intel Hyper-Threading Architectures. In Proceeding of IEEE Pacific-Rim Conference on Multimedia, 469-473 (2003)
6. Zhao Zhuo and Liang Ping.: A highly efficient parallel algorithm for H.264 video encoder. In Processing of 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 5, 489-492 (2006)
7. Jonghan Park and Soonhoi Ha.: Performance Analysis of Parallel Execution of H.264 Encoder on the Cell Processor. In Proceedings of the 2007 IEEE/ACM/IFIP Workshop on Embedded Systems for Real-Time Multimedia, 27-32, (2007)
8. Zhou X., Li E. Q., and Chen Y.-K.: Implementation of H.264 Decoder on General-purpose Processors with Media Instructions. In Proceeding of SPIE Conference on Image and Video Communications and Processing, 224-235, (2003)
9. Chuan-Yiu Lee, Yu-Cheng Lin, et al.: Multi-pass and frame parallel algorithms of motion estimation in H.264/AVC for generic GPU. In Proceeding of International Conference on Multimedia & Expo, 1603-1606, (2007)
10. Lin Chia-Chun, Lin Yu-Kun and Chang Tian-Sheuan.: PMRME: A Parallel Multi-Resolution Motion Estimation algorithm and architecture for HDTV sized H.264 video coding. In Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 2, 385-388, (2007)
11. Y.-W. Huang, T.-C. Chen, C.-H. Tsai, et al.: A 1.3TOPS H.264/AVC single-chip encoder for HDTV applications. In Proceedings of International Conference on Solid-State Circuits, 128-130. (2005)
12. D. M. Tullsen, S. J. Eggers, H. M. Levy.: Simultaneous Multithreading: Maximizing On-Chip Parallelism. In Proceeding of International Symposium on Computer Architecture, 392-403, (1995)
13. D. Marr, F. Binns, D. L. Hill, G. Hinton, D. A. Koufaty, J. A. Miller, and M. Upton.: Hyper-Threading Technology Micro-architecture and Architecture. http://developer.intel.com/technology/itj/2002/volume06issue01/art01_hyper/vol6iss1_art01.pdf (2002)

Haitao Wei is currently a Ph.D candidate at Huazhong University of Science and Technology in China. His research interests include parallel computing, compilation optimization, and parallel programming model.

Junqing Yu is the corresponding author of this paper. He received his PhD in Computer Science from Wuhan University in 2002. He is currently an Associate Professor at the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China. His research interests include digital media processing and retrieval, multi-core programming environment.

Haitao Wei, Junqing Yu, and Jiang Li

Jiang Li: was born in 1984. He got his master degree from Huazhong University of Science and Technology in 2008. His research interests include digital media processing and parallel computing.

Received: May 02, 2009; Accepted: August 13, 2009.

Robust moving object detection under complex background

DING Ying^{1,2}, LI Wen-hui¹, FAN Jing-tao², and
YANG Hua-min²

¹ School of Computer Science and Technology, Jilin University,
Changchun 130012, China
liwenhui2050@163.com

² School of Computer Science and Technology, Changchun University of Science
and Technology, Changchun 130022, China
{dingying, fanjingtao, yhm}@cust.edu.cn

Abstract. We present a novel method to robustly and efficiently detect moving object, even under the complexity background, such as illumination changes, long shadows etc. This work is distinguished by three key contributions. The first is the integration of the Local Binary Pattern texture measure which extends the moving object detection work for light illumination changing. The second is the introduction of HSI color space measure which removes shadows for the background subtraction. The third contribution is a novel fuzzy way using the Choquet integral which improves detection accuracy. The experiment results using several dataset videos show the robustness and effectiveness of the proposed method.

Keywords: moving object detection, Local Binary Pattern, HSI, Choquet integral.

1. Introduction

Background subtraction is often one of the first tasks in machine vision applications, making it a critical part of the system. The output of background subtraction is an input to a higher-level process that can be, for example, the tracking of an identified object. The performance of background subtraction depends mainly on the background modeling technique used to model the scene background. Especially natural scenes put many challenging demands on background modeling since they are usually dynamic in nature including illumination changes, swaying vegetation, rippling water, flickering monitors etc. A robust background modeling algorithm should also handle situations where new stationary objects are introduced to or old ones removed from the scene. Furthermore the shadows of the moving and scene objects can cause problems. Even in a static scene frame-to-frame changes can occur due to noise and camera jitter.

In this paper, we proposed a novel model for background maintenance and subtraction. The model aggregates color and texture features using fuzzy approach. The goal of the new model was to address all of the above-mentioned difficulties. Our contributions are: (1)Extend the background subtraction work for light illumination changes by integrating Local Binary Pattern (LBP) texture measure, (2)Remove shadows for the background subtraction by using HSI color measure, (3) Improve detection accuracy in a fuzzy way using the Choquet integral.

2. Related work

Different kinds of background model for detecting moving objects have been proposed in the literature, and some of them can be found to be robust to the challenges met in video sequence. These different methods are classified following the model used:

Basic Background Modeling (BBM): In this case, Background Representation is modeled using the average [1] or the median [2] or the histogram analysis over time [3]. Once the model is computed, the foreground detection is made as follows:

$$|f(I_t(x, y)) - f(B_t(x, y))| > Th \quad (1)$$

Otherwise, pixels are classified as background. Where This a constant threshold, $I_t(x, y)$ and $B_t(x, y)$ are respectively the current and the background images at time t , $f(x)$ is a feature value of x , such as intensity, gradient, etc.

Statistical Background Modeling (SBM): Background Representation is modeled using a single Gaussian [4-6] or a Mixture of Gaussians [7, 8] or a Kernel Density Estimation [9, 10]. Statistical variables are used in the foreground detection to classify the pixels as foreground or background. Recent SBM use Generalized Gaussian Mixture Modeling [11], Bayesian approaches [12, 13], Support Vector Regression learning approaches [14] or Codebook [15-17].

Background Estimation (BE): Background representation is estimated using a filter. For the foreground detection, any pixel of the current image that deviates significantly from its predicted value is declared foreground. This filter may be a Wiener filter [18], a Kalman filter [19] or a Tchebychev filter [20].

All these methods present the same following steps and issues: background modeling, background initialization, background maintenance, foreground detection, choice of the picture's element (pixel, a block or a cluster), choice of the features which characterize the picture's element (color features, edge features, stereo features, motion features and texture features). Often, these features are used separately and the most used is the color one. The combination of several measuring features can strengthen the pixel's classification as background or foreground. In a general way, the Choquet and Sugeno integrals have been successfully applied widely in classification problems [21], in decision making [22] and also in data modeling

[23] to aggregate different criteria. In the context of foreground detection, these integrals seem to be good model candidates for fusing different measures from different features. Each integral has its particularity. The Choquet integral requires to interpret the scale as a continuum and the Sugeno integral allows to work with an ordinal scale. Recently, Zhang and Xu [24] have used texture feature and color features to compute similarity measures between current and background pixels. Fida EL BAF[25] has fuzzy intensity and texture feature to foreground detection for infrared videos.

3. New approach

3.1. Approach overview

Moving object detection is based on a comparison between current and background images. In general, a simple subtraction is made between these two images to detect regions corresponding to moving object. So the choice of the features which characterize the pixel element is one of the most important steps. The other one is how to establish the comparison consists in defining a similarity measure between pixels in current and background images.

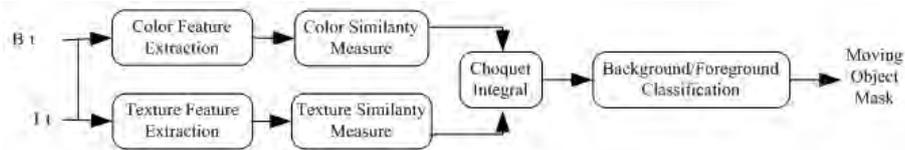


Fig. 1. Our approach overview

In this paper, we define a similarity measure between pixels in current and background images. In this case, pixels corresponding to background should be similar in the two images while pixels corresponding to foreground should not be similar. In Figure 1, the moving object detection process is presented in details. First, the color and texture features are extracted from the background image B_t and the current image I_t . The similarity measures are computed for each feature which is then aggregated by the Choquet integral. The Background/Foreground classification is finally made by threshold the Choquet integral's result. In the following subsections, we describe the rationale for selecting and fusing the set of the adopted features.

3.2. Color feature similarity measure

In order to remove the shadows' disturbance, pixels that could be part of a shadow have to be identified. RGB is the color space commonly acquired directly from a sensor or camera. HSI and YCbCr are closer to human interpretation of colors in the sense that brightness, for intensity, is separated from the base color. The best feature should decrease their sensitive to shadows. We choose HSI color space [26], and define the color features with HSI three components noted C_1 , C_2 and C_3 . Then, the color similarity measure $S_k(x, y)$ at the pixel (x, y) is computed as:

$$S_k(x, y) = 1 - \frac{|I_k(x,y) - B_k(x,y)|}{255} \quad (2)$$

Where $k \in \{1,2,3\}$ is one of three color features. $B(x, y)$ and $I(x, y)$ respectively represent the background and current frame at time t . Note that $S_k(x, y)$ is between 0 and 1. Furthermore, $S_k(x, y)$ is close to one if $B_k(x, y)$ and $I_k(x, y)$ are very similar.

3.3. Texture feature similarity measure

The proposed texture-based method for background subtraction is based on the Local Binary Pattern (LBP) texture measure. The LBP is a powerful means of texture description [27-29]. The operator labels the pixels of an image block by threshold the neighborhood of each pixel with the center value and considering the result as a binary number (LBP code):

$$LBP(x_c, y_c) = \sum_{i=0}^{K-1} f(p_i - p_c)2^i \quad (3)$$

Where p_c corresponds to the pixel value of the center pixel (x_c, y_c) , such as gray, intensity value etc. and p_i to the pixel values of the K neighborhood pixels. The function $f(x)$ is defined as follows:

$$f(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (4)$$

Then, for each pixel texture in the current image and the background image, the LBP code is less sensitive to illumination changes and is able to derive an accurate local texture difference measure [27]. To avoid light illumination changes' affect, here we define a texture similarity measure at pixel (x,y) between the current image and the background image as:

$$S_T(x, y) = 1 - \frac{|I_{LBP}(x,y) - B_{LBP}(x,y)|}{255} \quad (5)$$

Where $I_{LBP}(x,y)$ and $B_{LBP}(x,y)$ are respectively denotes the texture LBP code of pixel (x, y) in the background and current images. Note that $S_T(x,y)$ is between 0 and 1. Furthermore, $S_T(x,y)$ is close to one if $I_{LBP}(x,y)$ and $B_{LBP}(x,y)$

are very similar. In the false positive foreground areas caused by quick lighting changes, there are no texture changes between the current frame and the background. Hence, $S_T(x,y) \approx 1$. The foreground mask will be removed for the areas with $S_T(x,y) \geq T_s$. For this operation, we have chosen the choquet integrals.

3.4. Aggregation of Features by Choquet Integrals

Many fusion techniques can be used to fuse the color and the texture features. We present brief necessary concepts around fuzzy measures and the Choquet integrals [30].

Let λ be a fuzzy measure on a finite set X , and non-additive measure on a subset of X is any function $\mu: X \rightarrow [0, 1]$.

Definition 1 The Choquet integral of μ with respect to λ is defined by:

$$C_\lambda = \sum_{i=1}^n (\mu(x_{\sigma(i)}) - \mu(x_{\sigma(i-1)})) \lambda(A_{\sigma(i)}) \quad (6)$$

Where finite set $X = \{x_1, \dots, x_n\}$, and σ is a permutation of the indices such that $\mu_{\sigma(1)} \leq \dots \leq \mu_{\sigma(n)}$ and $A_{\sigma(i)} = \{\sigma(i), \dots, \sigma(n)\}$.

As defined above, the computed measures are obtained by dividing the feature values in background and current image with endpoints denoted by 0 and 1. For each pixel, color and texture similarity measures are computed as formula (2) (5) from the background and the current frame. We define the set of criteria $X = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ with $(\alpha_1, \alpha_2, \alpha_3) =$ three components color features of the chosen HSI color space and $\alpha_4 =$ texture feature $LBP(x, y)$.

With each criterion, we associate a fuzzy measure, for each α_i , let $\lambda(\alpha_i)$ be the degree of importance of the feature x_i in the decision whether pixel corresponds to background or foreground. Define $\lambda(\alpha_1) = \lambda(\{\alpha_1\})$, $\lambda(\alpha_2) = \lambda(\{\alpha_2\})$, $\lambda(\alpha_3) = \lambda(\{\alpha_3\})$, and $\lambda(\alpha_4) = \lambda(\{\alpha_4\})$ such that the higher the $\lambda(\alpha_i)$, the more important the corresponding criterion in the decision. To compute the fuzzy measure of the union of any two disjoint sets whose fuzzy measures are given, we use an operational version proposed by Sugeno [30] which called λ' -fuzzy measure. To avoid excessive notation, let denote this measure by λ' , where λ' is a parameter of the fuzzy measure used to describe an interaction between the criteria that are combined. Its value can be determined through the boundary condition, i.e. $\lambda'(X) = \lambda'(\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}) = 1$. The fuzzy density values over a given set $K \subseteq X$ are computed as:

$$\lambda'(K) = \frac{1}{\lambda'} \left[\prod_{x_i \in K} (1 + \lambda' \cdot \lambda(x_i)) - 1 \right] \quad (7)$$

The fuzzy function $\mu(\alpha_i)$ are defined in $[0, 1]$ so that, $\mu(\alpha_1) = S_1(x, y)$, $\mu(\alpha_2) = S_2(x, y)$, $\mu(\alpha_3) = S_3(x, y)$ and $\mu(\alpha_4) = S_T(x, y)$. To compute the value of Choquet integral for each pixel, we need firstly to rearrange the features α_i in the set X with respect to the order: $\mu(\alpha_1) \leq \mu(\alpha_2) \leq \mu(\alpha_3) \leq \mu(\alpha_4)$.

The pixel at position (x, y) is considered as foreground if its Choquet integral value is less than a certain threshold $T_{c,t}$, which denote the threshold at time instant t , as follows:

If $C_{\mu,t}(x, y) < T_{c,t}(x, y)$ then pixel (x, y) is foreground or moving object, else background.

4. Experimental results

The performance of the proposed method was evaluated using several video sequences. Both indoor and outdoor scenes were included. We have compared our method with the improved GMM modeling. Algorithms were implemented under Microsoft Visual C++ using the OpenCV library. The experimental results demonstrate the robustness of our algorithm in complex environments.

4.1. Experiments on indoor dataset



(a)background model (b)current frame (c)GMM detect result (d)our method detect result

Fig. 2. Moving object analysis of a indoor test sequence with a person come in

Fig. 2 compares a moving object detection result on the indoor test sequence from Wallflower [31], where a person is walking in a room, by GMM algorithm and our approach. Fig. 2a is the background model, and Fig. 2b is frame 650 (random choose) which after a person come in the office. The absolute color components change greatly with the illumination, even when no foreground object is present for the light changing. In Fig. 2c, large areas of false positive foreground were detected by the GMM method for light illumination change. As mentioned above, LBP is invariant to monotonic changes in gray scale. This makes it robust against illumination changes; Fig. 2d shows that our method successfully handles the light illumination changes by integrating texture information.

4.2. Experiments on outdoor dataset

Figure 3 shows the results of our algorithm for the outdoor test sequence, which contains changing environment and shadow. The original sequence has been taken from the PETS database [32] where several persons are walking in a subway station. The proposed algorithm successfully handles this situation. In HSI color space, the feature value changes of pixels in shadow region are very small, so most of the shadows can be removed by integrating HSI color information

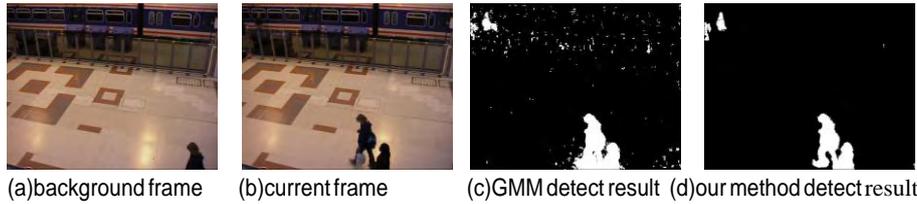


Fig. 3. Moving object analysis of an outdoor test sequence

4.3. Experiments on detection accuracy

To see the progression of the performance of each algorithm, we compute the true positive rate (TPR) and the false positive rate (FPR) as follows:

Let A be the ground truth point set and B be a detected region, the TPR and FPR can be defined as equation 8.

$$TPR = \frac{\sum_{(x,y) \in (A \cap B)} 1}{\sum_{(x,y) \in A} 1}, \quad FPR = \frac{\sum_{(x,y) \in (A \cap B)} 1}{\sum_{(x,y) \in A} 1} \quad (8)$$

Table 1 Detection results comparison

experiments	GMM Method		Our Method	
	TPR	FPR	TPR	FPR
1	1.00	0.056	1.00	0.00
2	0.958	0.012	0.968	0.003
3	0.945	0.011	0.947	0.003
4	0.945	0.011	0.951	0.004
5	0.957	0.012	0.960	0.004
6	0.952	0.012	0.968	0.004
7	0.940	0.012	0.949	0.004
8	0.948	0.021	0.950	0.001
9	0.943	0.020	0.959	0.008
10	0.989	0.237	0.985	0.002

Where (x, y) is a point in image. The TPR is the proportion of moving object pixels that were correctly classified among all positive samples. And FPR is the proportion of background pixels that were erroneously reported as being moving object pixels. In several experiments, the TPR and FPR result of different methods are compared as shown in Table 1. We have noticed that the TPR is very similar between these two methods, but the FPR of our method is quite lower than GMM's.

5. Conclusion

In this paper, we have presented a novel fuzzy background model for detecting moving objects from video frames. This method using Choquet integral for fuse color features and texture features. It chooses HSI color space instead of RGB, which remove most of the shadow, and aggregates LBP texture feature, which compute easily, to adapt the light illumination change. The proposed algorithm was tested against several standard benchmarks including both indoor and outdoor scenes. Further, the experiments results show that the proposed method is more robust and accurate.

6. References

1. B. Lee and M. Hedley: Background estimation for video surveillance. Image and Vision Computing New Zealand. (2002)
2. N. McFarlane and C. Schofield: Segmentation and tracking of piglets in images. British Machine Vision and Applications, Vol.8, No. 3, 187-193. (2005)
3. J. Zheng and Y. Wang: Extracting roadway background image: A mode based approach. Transportation Research Board, TRB, 82-88. (2006)
4. A. Francois and G. Medioni: Adaptive color background modeling for real-time segmentation of video stream. Proceedings of the International Conference on Imaging Science, Systems, and Technology, ICISST. (1999)
5. C. Wren and A. Azarbayejani. Pfister: Real-time tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.19, No. 7, 780–785. (1997)
6. M. Zhao, N. Li, and C. Chen: Robust automatic video object segmentation technique. IEEE International Conference on Image Processing ICIP. (2002)
7. C. Stauffer: Adaptive background mixture models for realtime tracking. Proceedings IEEE Conference on Computer Vision and Pattern Recognition, CVPR. (1999)
8. R. Tan, H. Huo, J. Qian, and T. Fang: Traffic video segmentation using adaptive-k gaussian mixture model. The International Workshop on Intelligent Computing, IWICPAS. (2006)
9. A. Elgammal and L. Davis: Non-parametric model for background subtraction. 6th European Conference on Computer Vision, ECCV. (2000)
10. C. Ianasi, V. Gui, C. Toma, and D. Pescaru: A fast algorithm for background tracking in video surveillance, using nonparametric kernel density estimation.

- Facta Universitatis, Series : Electronics and Energetics, Vol.18, No. 1, 127–144.(2005)
11. M. Allili, N. Bouguila, and D. Ziou: A robust video foreground segmentation by using generalized gaussian mixture modeling. Fourth Canadian Conference on Computer and Robot Vision, CRV. (2007)
 12. F. Porikli and O. Tuzel: Bayesian background modeling for foreground detection. ACM International Workshop on Video Surveillance and Sensor Networks, VSSN. (2005)
 13. Y. Sheikh and M. Shah: Bayesian modeling of dynamic scenes for object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI. Vol.27, No. 11, 1778–1792. (2005)
 14. J. Wang, G. Bebis, M. Nicolescu, M. Nicolescu, and R. Miller: Improving target detection by coupling it with tracking. Machine Vision and Application, MVA. (2008)
 15. A. Doshi and M. Trivedi: Hybrid cone-cylinder codebook model for foreground detection with shadow and highlight suppression. AVSS. (2006)
 16. K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis: Real-time foreground-background segmentation using codebook model. Real-Time Imaging. Vol.11, No. 3, 172-185. (2005)
 17. A. Leykin and R. Hammoud: Robust multi-pedestrian tracking in thermal-visible surveillance videos. OTCBVS. (2006)
 18. K. Toyama and J. Krumm: Wallflower: Principles and practice of background maintenance. International Conference on Computer Vision ICCV. (1999)
 19. S. Messelodi and C. Modena: A kalman filter based background updating algorithm robust to sharp illumination changes. International Conference on Image Analysis and Processing, ICIAP. (2005)
 20. R. Chang, T. Ghandi, and M. Trivedi: Vision modules for a multi sensory bridge monitoring approach. IEEE Conference on Intelligent Transportation Systems, ITS. (2004)
 21. H. Tahani and J. Keller: Information fusion in computer vision using the fuzzy integral. Transaction on Systems, Man, and Cybernetics, Vol.20, No. 3, 733–741. (1990)
 22. Y. Narukawa and T. Murofushi: Decision modelling using the choquet integral. Modeling Decisions for Artificial Intelligence, Vol.31, No. 31, 183–193. (2004)
 23. M. Sugeno and S. Kwon: A new approach to time series modeling with fuzzy measures and the choquet integral. 4th IEEE International Conference on Fuzzy Systems, ICFS. (1995)
 24. H. Zhang and D. Xu: Fusing color and texture features for background model. Third International Conference on Fuzzy Systems and Knowledge Discovery, Vol. 4223, No. 7, 887–893. (2006)
 25. Fida EL BAF: Thierry BOUWMANS and Bertrand VACHON: Fuzzy Foreground Detection for Infrared Videos. Computer Vision and Pattern Recognition Workshops, Anchorage, AK, 1-6. (2008)
 26. Ribeiro H, Gonzaga A: Hand image segmentation in video sequence by GMM: A comparative analysis. 19th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2006), Brazil, 357-364. (2006)
 27. M. Heikkilä, M. Pietikäinen and J. Heikkilä: A Texture-based Method for Detecting Moving Objects. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.28, No. 4, 657-666. (2006)
 28. T. Ojala, M. Pietikäinen, and D. Harwood: A comparative study of texture measures with classification based on feature distributions. Pattern Recognition, Vol. 29, No. 1, 51–59. (1996)

DING Ying, LI Wen-hui , FAN Jing-tao , and YANG Hua-min

29. T. Ojala etc.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 7, 971–987. (2002)
30. M. Grabisch, T. Murofushi, and M. Sugeno: Fuzzy Measures and Integrals. Theory and Applications (edited volume). Studies in Fuzziness. Physica Verlag. (2000)
31. Wallflower Dataset:
<http://research.microsoft.com/users/jckrumm/WallFlower/TestImages.htm>
32. PETS Dataset. <http://pets2006.net>.

Ding Ying is studying at Jilin University, she got Master Degree in 2007, at Changchun University of Science & Technology. Her main directions are image processing and computer simulation.

Li Wen-hui got his Ph.D. in 1994 from BREMEN University in Germany. He is currently a Professor of Computer Science at Jilin University.

Fan Jing-tao is studying at Changchun University of Science & Technology, he got Master Degree in 2007, at Changchun University of Science & Technology. His main directions are image processing and computer simulation.

Yang Hua-min got his Ph.D. in 2007 from Changchun University of Science & Technology in China. He is currently a Professor of Computer Science at Jilin University.

Received: April 13, 2009; Accepted: August 27, 2009.

A Novel Hierarchical Speech Emotion Recognition Method Based on Improved DDAGSVM*

Qi-rong Mao and Yong-zhao Zhan

Department of Computer Science and Communication Engineering of Jiangsu University, zhenjiang, Jiangsu Province, 212013, China
{mao_qr,yzzhan}@ujs.edu.cn

Abstract. In order to improve the recognition accuracy of speech emotion recognition, in this paper, a novel hierarchical method based on improved Decision Directed Acyclic Graph SVM (improved DDAGSVM) is proposed for speech emotion recognition. The improved DDAGSVM is constructed according to the confusion degrees of emotion pairs. In addition, a geodesic distance-based testing algorithm is proposed for the improved DDAGSVM to give the test samples differently distinguished many decision chances. Informative features and SVM optimized parameters used in each node of the improved DDAGSVM are gotten by Genetic Algorithm (GA) synchronously. On the Chinese Speech Emotion Database (CSED) and the Audio-Video Emotion Database (AVED) recorded by our workgroup, the recognition experiment results reveal that, compared with multi-SVM, binary decision tree and traditional DDAGSVM, the improved DDAGSVM has the higher recognition accuracy with few selected informative features and moderate time for 7 emotions.

Keywords: Speech Emotion Recognition, Improved DDAGSVM, Hierarchical Recognition Method, Confusion Degree, Geodesic Distance.

1. Introduction

Speech emotion recognition plays an important role in affective computing. Up to now, many pattern recognition methods have been used in the speech emotion recognition. For example, ANN[1], HMM[2-3], GMM[4], Gaussian supervector with SVM [5]. In paper [6] and [7], visual and audio signals were fused to detect emotion by using many recognition methods. Despite many promising results have been gotten in the past, it is not satisfying. Research results have shown that contributions of different features to an emotion are different[8], and the separabilities between different emotions are different[9]. This finding implies that it maybe improve the recognition accuracy of speech emotion to use effective features to recognize different emotions by the means

* Supported by the National Natural Science Foundation of China under Grant No.60673190, and the University Science Research Project of Jiangsu Province No.09KJB520002.

of hierarchical classification method. But up to now, in many of speech emotion recognition methods, classifiers are used to classify all emotions at the same layer by using the same feature subset.

SVM has good classification ability, and it is a kind of 2-class classifier. So it is very easy to use SVM to construct the hierarchical classification method. Binary decision tree is one kind of hierarchical classification method by using SVM[10]. It only needs $N-1$ SVMs when N emotions need to be classified. But in each node of SVM-based binary decision tree, classifier is 1-vs- r SVM, and the training samples of it are unbalance in most cases. Furthermore, the training of 1-vs- r SVM is more time-consuming than 1-vs-1 SVM.

SVM-based Decision Directed Acyclic Graph (DDAGSVM) is another kind of hierarchical classification method by using SVM[11]. In this method, 1-vs-1 SVM is used as the classification function of each node. A 1-vs-1 SVM can only exclude one class from consideration classes. For a problem with N classes, $N-1$ decision nodes will be evaluated in order to derive an answer, and there are N layers in the DDAG. And $N*(N-1)/2$ 1-vs-1 SVMs are trained in DDAG. The path taken through the DDAG SVM from root node to the last node is known as the evaluation path. Research result in paper [11] shows that DDAGSVM is amenable to a VC-style bound of generalization error, and that it yields comparable accuracy and memory usage to the standard 1-v- r SVM. But in the traditional DDAGSVM, the choice of the class order is arbitrary. The unclassifiable regions are assigned to the classes associated with the leaf nodes. In paper [12], an optimizing DDAGSVM is proposed to improve the generalization ability of DDAG. This research work enlightens us.

In another aspect, there are many feature selection methods proposed for the speech emotion recognition recently, which can be broadly classified as random or nonrandom selection[13-16]. The selected features by the random selection methods can get higher recognition accuracy on the classifier employed. Moreover, informative features selection and SVM parameter optimization can be finished synchronously by Genetic Algorithm (GA). Thus, GA is designed to select features and optimize parameters in this paper.

In this study, an improved DDAGSVM is proposed to reduce error classification cumulation of DDAGSVM, and then a novel hierarchical speech emotion recognition method is put forward by using the improved DDAGSVM. In addition, GA is used to select informative features for each class pair and optimize SVM parameters in each node.

Section 2 describes the confusion degrees of emotion pairs. Hierarchical speech emotion recognition method based on improved DDAGSVM is presented in section 3. Experiment results on the Chinese Speech Emotion Database (CSED) and the Audio-Video Emotion Database (AVED) for 7 emotions are illustrated in section 4. Section 5 draws a conclusion.

2. Confusion degrees of emotion pairs

Definition 1 (Confusion degree between two emotions): It is assumed that E_i and E_j denote the i^{th} emotion and the j^{th} emotion respectively. Then, confusion degree between E_i and E_j is the mean of the probability that the samples labeled E_i are classified into E_j and the probability that the samples labeled E_j are classified into E_i . It is denoted by I_{ij} , and it is defined as Eq. (1).

$$I_{ij} = \frac{p(r=i | x \in E_j) + p(r=j | x \in E_i)}{2} \quad (1)$$

In Eq.(1), x is a test sample, and r is the classification result. We can see clearly that, the greater the confusion degree between two emotions is, more difficultly the two emotions are differentiated. The confusion degrees of emotion pairs can be gotten by confusion matrix.

3. Improved DDAGSVM

In this paper, to improve the error classification cumulation of the DDAGSVM, an improved DDAGSVM is proposed. Compared with the traditional DDAGSVM, improved DDAGSVM has two improvements. One is the constructing method of DDAGSVM, and the other is the geodesic distance-based testing algorithm for the test samples differently distinguished. In the improved DDAGSVM, emotion pairs having smaller confusion degrees are classified in the upper nodes, and emotion pairs with greater confusion degrees are differentiated in the lower nodes. The improved DDAGSVM is constructed with many evaluation paths. According to confusion degrees of emotion pairs, an improved DDAGSVM can be set up.

In DDAGSVM, if the recognition accuracy of the test sample differently distinguished is increased, the classification ability of DDAGSVM can be improved. In the following, 1-vs-1 SVM is analyzed to find the samples which may be classified in error.

Eq. (2) is the equation of the hyperplane of SVM[17]. It makes the margin between two classes largest when $f(x)=0$. The hyperplane is denoted as H_0 . In Eq.(2), $K(x_i, x)$ denotes kernel function, and x_i ($i=1,2,3...m$) denotes m -dimensional inputs. y_i denotes the class label, and $y_i=1$ for class 1 and $y_i=-1$ for class 2. Two parallel plane of H_0 are called as insulation plane, and they are denoted as H_1 and H_2 respectively. The equation of plane H_1 is $f(x)=1$, and the equation of plane H_2 is $f(x)=-1$. They are shown in Fig.1. For the position of a test sample x_0 , there are three possible situations.

$$f(x) = \sum_{i=1}^m y_i \alpha_i^* k(x_i, x) + b^* \quad (2)$$

Situation 1: x_0 locates on or above the plane H_1 , namely, $f(x_0) \geq 1$;

Situation 2: x_0 locates on or under the plane H_2 , namely, $f(x_0) \leq -1$;

Situation 3: x_0 locates in the margin between the plane H_1 and the plane H_2 , namely, $-1 < f(x_0) < 1$;

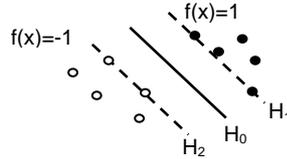


Fig. 1. Largest margin of SVM

If situation 1 or situation 2 happens, x_0 is classified as class 1 or class 2 correctly. But if situation 3 happens, x_0 may be classified by error in all probability. Therefore, the key to reduce error classification cumulation of DDAGSVM is whether the test samples in situation 3 can be classified correctly. In this paper, except the leaf nodes in the improved DDAGSVM, a new testing algorithm for each node is proposed as follows.

Algorithm 1. Geodesic distance-based testing algorithm for each node

It is assumed that T_{sv1} and T_{sv2} denote the support vector set of class 1 and class 2 respectively. It is also assumed that there are n_1 support vectors in T_{sv1} and n_2 support vectors in T_{sv2} . x_0 denotes a test sample. $f(x_0)$ denotes the distance between x_0 and the hyperplane H_0 of SVM. $GD(x_0, s_i)^1$ denotes the geodesic distance between x_0 and the i^{th} support vector in T_{sv1} , and $GD(x_0, s_i)^2$ denotes the geodesic distance between x_0 and i^{th} support vector in T_{sv2} . GD_{mean}^1 and GD_{mean}^2 denote the average distances between x_0 and support vectors of class 1 and class 2 respectively.

x_0 is classified by SVM, and $f(x_0)$ can be gotten;

The label of x_0 recognized by SVM is saved as L_0 ;

If $(f(x_0) \leq -1$ or $f(x_0) \geq 1)$

{ x_0 is classified as the class denoted by L_0 ;

else if $(-1 < f(x_0) < 1)$

{ Compute $GD(x_0, s_i)^1$ and $GD(x_0, s_i)^2$ according to Algorithm 2;

$$GD_{mean}^1 = \frac{1}{n_1} \sum_{i=1}^{n_1} GD(x_0, s_i)^1;$$

$$GD_{mean}^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} GD(x_0, s_i)^2;$$

If $((|GD_{mean}^1 - GD_{mean}^2| < \delta)$ and $(L_0$ is class 1))

{ x_0 is classified as class 1;}

else if $((|GD_{mean}^1 - GD_{mean}^2| > \delta)$ and $(L_0$ is class 2))

{ x_0 is classified as class 2;}

else if $((|GD_{mean}^1 - GD_{mean}^2| < \delta)$ and $(L_0$ is class 2))

{ if $(|GD_{mean}^1 - GD_{mean}^2| > \delta)$

{ x_0 is classified as class 1;}

else if $(|GD_{mean}^1 - GD_{mean}^2| \leq \delta)$

{ x_0 is sent to both the left and the right node to be classified in further. The

class label of x_0 is decided by the membership degree sequences of evaluation path, which is defined in the last paragraph of this section.}}
 else if ((($GD_{mean}^1 - GD_{mean}^2$) > 0) and (L_0 is class 1))
 { if ($|GD_{mean}^1 - GD_{mean}^2| > \delta$)
 { x_0 is classified as class 2;}
 else if ($|GD_{mean}^1 - GD_{mean}^2| \leq \delta$)
 { x_0 is sent to both left and right node to be classified in further;}}

In Algorithm1, there are two reasons to use the average geodesic distance between a test sample and support vectors of one class to reflect the distance between the test sample and the class. One reason is that the geodesic distance between samples can reflect the factual distance between samples in feature space[18]. The other reason is that support vectors of one class can be regarded as the representative samples of this class. It consumes less time than calculating distance between a test sample and all the training samples of one class. The calculating method of geodesic distance between sample x_0 and a support vector s_i is shown as Algorithm 2.

Algorithm 2. Calculating method of the geodesic distance $GD(x_0, s_i)$ between the sample x_0 and the support vector s_i .

Input: x_0 and s_i

Output: the geodesic distance $GD(x_0, s_i)$ between the sample x_0 and the support vector s_i .

- 1) T_{sv1} and T_{sv2} are merged into one set T , and x_0 is inserted into T . The elements in T are denoted as v .
- 2) Calculate the euclidean distances of every two samples in T .
- 3) Calculate k neighbors of each sample in T . The k neighbors of v_i is denoted as $\Omega(v_i)$.
- 4) Create the neighbor matrix G according to Eq.(3) in the following.

$$G(i, j) = \begin{cases} d(v_i, v_j) & v_j \in \Omega(v_i) \\ \infty & v_j \notin \Omega(v_i) \end{cases} \quad (3)$$

Where, $G(i, j)$ denotes the element of i^{th} line and j^{th} column in matrix G . $d(v_i, v_j)$ denotes the euclidean distance between v_i and v_j .

- 5) Estimate the shortest distance $GD(v_i, v_j)$ between v_i and v_j according to Floyd Algorithm.

$$GD(v_i, v_j) = \min(GD(v_i, v_j), GD(v_i, v_k) + GD(v_k, v_j)) \quad (4)$$

When v_i is x_0 , and v_j is s_i , $GD(v_i, v_j)$ is the geodesic distance between x_0 and s_i . In Algorithm 1, when a test sample is sent to both the left and the right node to be classified in further, which evaluation path of the test sample is reliable? In this paper, the membership degree sequence of evaluation path is defined to judge the most reliable evaluation path. Here, if it happens in l^{th} node that the test sample is sent to both the left and the right node according to Algorithm 1, the evaluation path is checked from $(l+1)^{th}$ node to the last node. Since each node eliminates one class from the list in the evaluation path of improved DDAGSVM, the membership degrees of the test sample x_0 not belonging to emotion class 1 and emotion class 2 in the node l are defined as

Eq.(5). They are denoted by λ_i^1 and λ_i^2 respectively.

$$\lambda_i^1 = \begin{cases} 0 & f(x_0) \geq 1 \\ \frac{1}{GD_{mean}^2} & -1 < f(x_0) < 1 \\ 1 & f(x_0) \leq -1 \end{cases} \quad \lambda_i^2 = \begin{cases} 1 & f(x_0) \geq 1 \\ \frac{1}{GD_{mean}^1} & -1 < f(x_0) < 1 \\ 0 & f(x_0) \leq -1 \end{cases} . \quad (5)$$

When the membership degrees of x_0 in each node traveled are gotten, these membership degrees make up of membership degree sequences for evaluation paths. It is assumed that there are two sequences of two evaluation paths P_j and P_k , and they are denoted as $q_j = (\lambda_{(l+1,j)}^i, \lambda_{(l+2,j)}^i, \lambda_{(l+3,j)}^i, \dots, \lambda_{(N-1,j)}^i)$ and $q_k = (\lambda_{(l+1,k)}^i, \lambda_{(l+2,k)}^i, \lambda_{(l+3,k)}^i, \dots, \lambda_{(N-1,k)}^i)$ ($i \in \{E_1, E_2, \dots, E_N\}$). Lexicographic order is used as the rule to compare two membership degree sequences. The test sample x_0 is classified as the emotion class of the leaf node of P_j if and only if $q_j > q_k$ according to the lexicographic order.

4. Experiment Results and Discussions

4.1. Selecting informative features and optimizing SVM parameters

In this paper, two emotion speech corpora are used for experiments. One corpus is CSED [10], and the other is AVED [19]. These two emotion corpora are recorded by our workgroup. Each database includes short utterances covering 7 emotions. In this paper, 132 dimensional original features are extracted from the aspects of the voice of quality, energy, pitch, formant frequency, MFCCs (Mel Frequency Cepstral Coefficients), Mel Frequency energy Dynamic Coefficient and multi-fractal [19-21].

By using GA, we can select informative features for each emotion pair and optimize parameters for the corresponding SVM. Both the recognition accuracy and the number of selected features are taken into consideration in the fitness function, which is defined by Eq.(6).

$$fitness = w_r * SVM_accuracy + 10 * w_n * features_num^{-1} . \quad (6)$$

In Eq.(6), $SVM_accuracy$ denotes SVM recognition accuracy, w_r denotes the SVM recognition accuracy weight, w_n denotes the features number weight, and $features_num$ denotes the number of the features selected. The constant 10 in the second item of Eq.(6) is to make the two items keep balance. According to the fitness function of Eq. (6), w_r and w_n can influence the experiment result. We defined $w_r = 0.8$ and $w_n = 0.2$ in this paper according to experiment. The best chromosome is obtained when the termination criteria is satisfying. Both the selected informative features and the pair of (C, σ) for each SVM can be gotten by GA. Here, $C \in [2^{-5}, 500]$, $\sigma \in [2^{-10}, 100]$. The number of

the united informative features of all emotion pairs is 46. That is to say, only 46 features need to be extracted from samples before testing. It improves the efficiency of the improved DDAGSVM.

4.2. Constructing Improved DDAGSVM

Before confusion degrees of emotion pairs are calculated according to Eq.(1), confusion matrix of 7 emotions must be gotten. It is gotten by using multi-SVMs. Feature subset and SVM parameters used in this step are gotten by GA. The confusion matrix on CESD is shown in Table 1. According to the confusion matrix, confusion degrees of emotion pairs can be gotten, and then the improved DDAGSVM is set up. The part of it is shown in Fig.2.

Table 1. Confusion matrix by multi-SVMs on CESD

%	Happiness	Sadness	Surprise	Anger	Fear	Disgust	Neutral
happiness	83.3	0	6.7	6.7	3.3	0	0
Sadness	0	80	0	0	0	13.3	6.7
Surprise	6.7	0	83.3	3.3	3.3	0	3.3
Anger	0	0	10	90	0	0	0
Fear	0	3.3	0	0	86.7	3.3	6.7
Disgust	3.3	0	0	0	0	90	6.7
Neutral	3.3	6.7	0	0	3.3	20	66.7

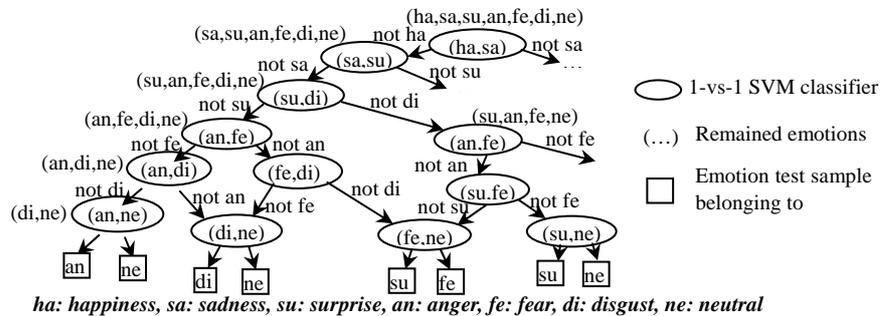


Fig.2. Part of the improved DDAGSVM for 7 emotions

4.3. Recognition experiments

For the improved DDAGSVM, the value of δ in Algorithm 1 and the value of k in Algorithm 2 are important, and they are difficult to choose. If the value of δ is adjusted to greater, the recognition accuracy may be improved, but the number of the test samples that need to be classified in many evaluation paths will increase. It will consume much time to test. On the contrary, if the value of δ is adjusted to smaller, the recognition accuracy may be decrease. In this paper, we compromised between the recognition accuracy and the test time.

50 is chosen as the value of δ according to the experiment. Moreover, according to the experiment, 6 is chosen as the value of k .

The hierarchical speech emotion recognition method based on improved DDAGSVM proposed in this paper was evaluated on two different emotion speech databases CSED and AVED. On each database, we evaluated five methods for 7 emotions:

- Multi-SVM. 21 1-vs-1 SVMs of it are executed in the same layer.
- SVM-based Binary Decision Tree(SVM-based BDT). In this decision tree, the emotions having the better clustering performance are distinguished in the upper layer, and the emotions having the worse clustering performance are distinguished in the lower layer [10].
- Traditional DDAGSVM. The choice of the class order of this method is arbitrary.
- Improved DDAGSVM not with the Geodesic distance-based testing algorithm. In this method, the DDAGSVM is constructed according to the confusion degrees of emotion pairs. The training and testing processes of this method are the same with those of the traditional DDAGSVM.
- Improved DDAGSVM with the Geodesic distance-based testing algorithm proposed in this paper (improved DDAGSVM with GD test).

In the method mentioned above, both the features and the parameters used by each SVM are selected or optimized by GA. The recognition accuracies of these five methods on the two speech emotion databases are listed in Table 2. Table 3 list the number of SVMs needed to train and to test. The mean training time and the testing time of one sample needed by the five classification methods are also listed.

Table 2. Average recognition accuracy on the two databases

Database	Classification methods	Recognition accuracy(%)
CSED	Multi-SVM	77.5
	SVM-based BDT	78.8
	Traditional DDAGSVM	78.9
	Improved DDAGSVM not with GD test	80
	Improved DDAGSVM with GD test	82.7
BDESAVED	Multi-SVM	74.8
	SVM-based BDT	76.9
	Traditional DDAGSVM	76.8
	Improved DDAGSVM not with GD test	78.9
	Improved DDAGSVM with GD test	80.5

From the results in Table 2, we can see clearly that the hierarchical methods have better recognition accuracy than multi-SVM. It proves that the hierarchical methods can improve the recognition accuracy of speech emotion recognition. Compared with the other four methods, the improved DDAGSVM with GD test proposed in this paper has the highest recognition accuracy on the two databases. And the improved DDAGSVM not with GD test has better recognition than the traditional DDAGSVM, SVM-based BDT and multi-SVM. These experiment results indicate that the construction method of DDAGSVM and the geodesic distance-based testing algorithm proposed in this paper can

improve the classification ability of DDAGSVM. In the aspect of time consumed, the results in Table 3 illustrate that SVM-based BDT needs the longest time to train. It is because that the classifier in binary decision tree is 1-vs-r SVM, the training of which will consume much time. The testing time consumed by the improved DDAGSVM not with GD test is the same with that of the traditional DDAGSVM. But the improved DDAGSVM with GD test has longer testing time than the two methods, because the calculation of geodesic distance and the membership degree sequence will consume much time when the test sample is distinguished differently. Although the improved DDAGSVM with GD test will consume much time in testing, the increasing testing time is little. Compared with the least testing time 0.026s, the average testing time of the improved DDAGSVM with GD test only increases 0.006s.

Table 3. Consumed time by the five methods

Classification method	The number of SVM needed to train	The number of SVM needed to test	Training time(s)	Average testing time of one test sample(s)	Maximum testing time of one sample (s)
Multi-SVM	$N*(N-1)/2$ (1-vs-1)	$N*(N-1)/2$ (1-vs-1)	15	0.035	0.035
SVM-based BDT	N-1 (1-vs-r)	$<(N-1)$ (1-vs-r)	26.6	0.03	0.038
Traditional DDAGSVM	$N*(N-1)/2$ (1-vs-1)	N-1 (1-vs-1)	15	0.026	0.026
Improved DDAGSVM not with GD test	$N*(N-1)/2$ (1-vs-1)	N-1 (1-vs-1)	15	0.026	0.026
Improved DDAGSVM with GD test	$N*(N-1)/2$ (1-vs-1)	$>N-1$ (1-vs-1)	15	0.032	0.041

5. Conclusions

In this paper, a novel hierarchical method based on improved DDAGSVM was proposed for the speech emotion recognition. Improved DDAGSVM is set up according to the confusion degrees of emotion pairs. In addition, a geodesic distance-based testing algorithm is proposed to give the test samples differently distinguished many decision chances. The informative features for each emotion pair are selected by GA while the SVM parameters are optimized by GA. Before testing, only 46 features are extracted from testing samples. This makes training and testing of the improved DDAGSVM more quickly. The classification experiments have been done on two speech emotion databases CSED and AVED. The experiment results reveal that, compared with the other four methods, the improved DDAGSVM with the Geodesic distance-based testing algorithm has the highest recognition accuracy. The disadvantage of it is that the testing time is longer. But the increasing extent of it is small. It is still fit for the real-time applications. This is worth to be studied in the further work. The method proposed in this paper can be applied in other pattern recognition fields.

6. References

1. Bhatti M. W., Wang Y., Guan L.: A Neural Network Approach for Human Emotion Recognition in Speech. In Conf. ISCAS. Vancouver, BC, Canada, 181-184. (2004)
2. Nwe T. L., Foo S. W., Silva L. C. D.: Speech Emotion Recognition Using Hidden Markov Models. *Speech Communication*, Vol. 41, No. 4, 603-623. (2003)
3. Schuller B., Rigoll G., Lang M.: Hidden Markov Model-Based Speech Emotion Recognition. In IEEE Int. Conf. ICASSP. New York, USA, 1-4. (2003)
4. Moataz M. H., Mohamed S., Fakhri K.: Speech Emotion Recognition Using Gaussian Mixture Vector Autoregressive Models. In IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP). Beijing, China, IV-957-960. (2007)
5. Hu H., Xu M. X., Wu W.: GMM Supervector Based SVM with Spectral Features for Speech Emotion Recognition. In Conf. of ICASSP. Beijing, China, 413-416. (2007)
6. Wang Y. J., Guan L.: Recognizing Human Emotional State From Audiovisual Signals. *IEEE TRANSACTIONS on Multimedia*, Vol.10, No.4, 659-667. (2008)
7. Zeng Z. H., Tu J. L., Brian M. P., Thomas S. H.: Audio-Visual Affective Expression Recognition Through Multistream Fused HMM. *IEEE Transaction on Multimedia*, Vol.10, No.4, 570-577. (2008)
8. Lin Y. L., Wei G., Yang K. C.: A Survey of Emotion Recognition in Speech. *Journal of Circuits and Systems*, Vol.12, No.1, 90-98 (in Chinese). (2007)
9. Dimitrios V., Constantine K.: Emotional Speech Recognition: Resources, Features, and Methods. *Speech Communication*, Vol. 48, 1162-1181. (2006)
10. Mao Q. R., Wang X. J., Zhan Y. Z., Pan Z. G.: Speech Emotion Recognition Method Based on Selective Features and Decision Binary Tree. *Journal of Computational Information Systems*, Vol. 4, No.4, 1795-1801. (2008)
11. John C. P., Nello C., John S. T.: Large Margin DAGs for Multiclass Classification, S.A. Solla, T.K. Leen and K.-R. Müller (eds.), MIT Press, 547-553. (2000)
12. Takahashi, F., Abe, S.: Optimizing Directed Acyclic Graph Support Vector Machines. In Proc. of ANN in Pattern Recognition, 166-170. (September 2003)
13. Huang C. H., Wang C. J.: A GA-based Feature Selection and Parameters Optimization for Support Vector Machines. *Expert Systems with Applications*, Vol. 31, No. 2, 231-240. (2006)
14. Ververidis D., Kotropoulos C.: Fast and Accurate Sequential Floating Forward Feature Selection with the Bayes Classifier Applied to Speech Emotion Recognition. *Signal Processing*, Vol.88, 2956-2970. (2008)
15. Altun H., Polat G.: Boosting Selection of Speech Related Features to Improve Performance of Multi-class SVMs in Emotion Detection. *Expert Systems with Applications*, Vol. 36, No. 4, 8197- 8203. (2009)
16. Wang X. J., Mao Q. R., Zhan Y. Z.: Speech emotion feature selection method based on contribution analysis algorithm of neural network. *AIP Conference proceedings*, Vol.1060, 336-339. (2007)
17. Hsu C. W. and Lin C. J.: A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Transaction on Neural Networks*, Vol. 13, No. 2, 415-425.(2002)
18. You M. Y., Chun C., Bu J. J., Tao J. H.: Analysis on Nonlinear Manifold. In proc. of ICPR'06, Vol.3, 91-94. (2006)
19. Mao Q. R.: Research on the feature extraction and the recognition method of speech emotion. Ph. D. thesis. Zhenjiang, Jiangsu University, 27-28 (in Chinese). (2009)
20. Roddy C., Randolph R. C.: Describing the Emotional Classes That are Expressed in Speech. *Speech Communication*, Vol. 40, 5-32. (2003).

A Novel Hierarchical Speech Emotion Recognition Method Based on Improved
DDAGSVM

21. Zhan Y. Z., Cao P.: Research and Implementation of Emotional Feature Extraction and Recognition in Speech Signal, Journal of Jiangsu University (National Science Edition), Vol. 26, No.1, 72-75(in Chinese). (2005)
22. Lin Y. L.: Research on Speech Emotion Recognition. Ph.D. Thesis. Guangzhou, South China University of Technology, 51-57, (in Chinese). (2006)

Qirong Mao got her M.S. degree in Computer Science from Jiangsu University in China, in 2002. She is currently an associate professor at the Department of Computer Science and Communication Engineering in Jiangsu University at Zhenjiang (China). Now she is working on the affective computing for the Ph.D.. Her current research interests include Human-Computer Interaction (HCI) and Computer Supported Collaborative Work (CSCW).

Yongzhao Zhan is a Professor of Computer Science at Jiangsu University, Zhenjiang, China. He got his Ph.D. in Computer Science from Nanjing University, China, in 2000. His current research focuses on Human-Computer Interaction (HCI), Computer Supported Collaborative Work(CSCW) and the security of wireless network.

Received: May 05, 2009; Accepted: September 29, 2009.

Research on Optimizing the Fault Diagnosis Strategy of Complex Electronic Equipment

WANG Hong-xia¹, YE Xiao-hui¹, and WANG Liang²

¹ Naval University of Engineering, Electronic Engineering College, Electronic Engineering Department, Wuhan 430033, China,
{ mimiconan }@ sina.com

² Navy 705 Factory; Zhanjiang 524006, China

Abstract. Diagnosis strategy is a testing sequence of the fault detection and isolation. For the distribution of the electronic equipment, a feasible engineering maintenance method is put forward based on the questions of test point selection and diagnosis strategy. The concepts of local diagnosis strategy and global diagnosis strategy are introduced. From which the local optimal diagnosis strategy is determined when the local optimal test points have been introduced by using the test information entropy, furthermore, the global optimal diagnosis strategy is determined by coalescing the local optimal diagnosis strategies. At last, the validity of the method is illustrated by an example from which the conclusion can be drawn that it is an optimal diagnosis strategy and the complexity of computation can be reduced.

Keywords: electronic equipments, dependency matrix, optimizing strategy.

1. Introduction

The modern complex Electronic equipments are made up by some sub-systems which are different from structure to function. The sub-systems are linked by LAN or BUS to achieve the respective functions. Though the functions and the effects of sub-systems are different, the complexity and distribution are similar to solve the problems in the fault diagnosis. Wang [1] has presented a method for the fault diagnosis modeling of complex electronic equipments, which could achieve the FDR (Fault Detection Ratio) and FIR (Fault Isolation Ratio) effectively. The objective of the FDI (Fault Detection and Isolation) is to provide an effective and feasible fault detection sequence for the maintenance personal to locate the fault to the LRU (Line Replaceable Unit) in the maintenance task. So a convenient and feasible fault diagnosis strategy is very important for the maintenance of the equipments.

A multi-signal dependency model is akin to overlaying a set of (single-signal) dependency models on the structural model, and, hence, the model corresponds closely to the schematics of the system. The system fault diagnosis strategy corresponding closely to the schematics of the system is

the optimal strategy based on the test point selection and dependency matrix which use the relationship between the components and tests. Tian[2] gives an example to show that test point selection can influence the selection of the fault diagnosis strategy. The objective of the fault diagnosis strategy is to isolate fault with the minimal time and expense. Jing[3] has used the method of information entropy to get the fault diagnosis strategy which used the information of test outcome, test cost and failure rate to get the lowest test cost. But there are thousands of test points, and the size of information entropy of test point which test the same function is not adjacent, the valid diagnosis strategy cannot achieved by using the Jing's methods. Therefore, the optimizing diagnosis strategy is presented.

This paper is organized as follows. In section 2, the basic concepts are presented. In section 3, the structure and function of the electronic equipments are decomposed and then the system global diagnostic strategy by coalescent local diagnostic strategy is generated. In section 4, we apply the optimizing diagnosis strategy to the some electronic equipment, and the simulation results are presented. Finally, the paper concludes with a summary in section 5.

This paper decomposes the structure and function of the electronic equipments, then, generates the system global diagnostic strategy by coalescent local diagnostic strategy.

2. Preliminaries

2.1. Dependency Matrix

The directed graph model captures the first order cause-effect dependencies such as A affects B and B affects C. The global dependencies, such as A affects C, are inferred by the reachability analysis algorithms. Specifically, we need to ascertain which of the failure sources can be observed from each of the tests of the test points, thus enabling us to compute the dependency matrix.

Similarly, a test point may have multiple tests associated with it. In this case, if there are m components and n tests, the dependency matrix (D-matrix) is of size $(m+1) \times n$, where $d_{ij}=1$, if the fault component c_i can be detected by the test t_j ; else $d_{ij}=0$. The D-matrix summarizes the diagnostic information of the system and all analysis is performed using this matrix.

2.2. Test Point Optimal Selection

The fault detection is to judge whether fault exists on UUT (Unit Under Test), therefore we should choose the signals which associate more composing unit

to test in order that we can use the fewest test to determine whether the fault exists. Suppose that the refined dependency matrix is $D=[d_{ij}]_{m \times n}$, we can use the following formula to calculate WFD_j (the fault detect weight of the test j),

$$WFD_j = \sum_{i=1}^m d_{ij}$$

That is, after calculating WFD_j weight of each test, choose the test with the maximum weight as priority test when fault detecting. Dividing the dependency matrix into two parts with corresponding column matrix, we gain two sub-matrixes. According to the test result, we choose the test which has the biggest weight in the sub-matrix as the second test, then repeat above-mentioned process until there is not column matrix any more.

Fault isolation is a process which isolates fault to the LRU, the process of test point selection for fault isolation is similar to the process for fault detection. Please refer to the reference[4].

Making diagnosing strategy is based on test's optimal selecting outcome. When the state of UUT is unknown, we select optimal test for fault detection first, if it passes, continue to do optimal test; if it doesn't pass, switch to fault isolation process. Repeating above-mentioned process in turn until the fault detection and isolation finish, we will get a fault tree finally.

3. Optimizing Fault Diagnosis Strategy

3.1. Method of Decomposition for Electronic Equipment Diagnosis

Modern electronic equipment has hierarchical structure, the characteristic of system structure bring the hierarchical faults. Hierarchical fault strategy was generated by adopting hierarchical fault diagnosis, ultimate purpose of which is to isolate fault to LRU effectively, let the technician use these results to isolate the abnormal components. Therefore decomposition system is important.

Systematic decomposition principle is that the subsystem function is clear and the coupling between the function is little as much as possible, so that common component was called little. Generally speaking, Structure decomposition was used in coarse granularity, while function decomposition was used in fine granularity, which was coincide with the thought of designing the embedded system.

In this paper the structure decomposition (Fig. 1.), the function decomposition (Fig. 2.) is carrying out to decompose electronic equipments. Structure decomposition is to division system on structure until the lowest component, based on which we can precisely locate the fault to the physics place.

The function decomposition is to division system function to the sub-function, from which we can get the basic function. Based on above decomposition, the diagnosis strategy algorithm can only search space which contains some components relevant to that function.

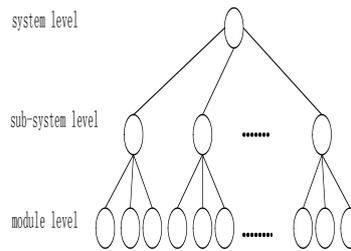


Fig. 1. Structure decomposition

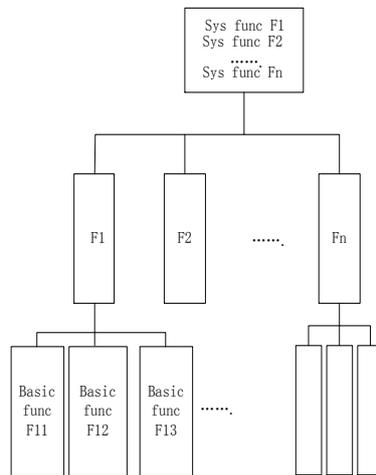


Fig. 2. Function decomposition

3.2. Global Diagnosis Strategies

Hypothesis 1: Regular signal behavior is dependent on all components normally, if any component behavior is abnormal then the signal is abnormal.

Hypothesis 2: 1) the signal of relevant to the function is only limited to the correlative component. 2) the relationship between different signals is uncorrelated.

Definition 1: Local diagnosis strategy (LDS): A diagnosis and isolation sequence for all the components to realize certain function.

Definition 2: Global diagnosis strategy (GDS): A diagnosis and isolation sequence for a certain system

Local diagnosis strategy is a binary tree: the left branch indicates that the test pass, the new binary tree was deployed by selecting the optimum test point for the fault detection, until all the tests pass, which indicates that system is fault-free; the right branch indicates that the test fails, the new binary tree was deployed by selecting the optimum test point for the fault isolation, until all the faults can be isolated. Test point for the fault detection and the fault isolation in the local diagnosis strategy are columns of the local dependency matrix.

In this paper, components partition is two parts: private component and common component; while the private component is contained only in a function, the common component contains in at least two functions. Private component sets may be further divided into sub private component belonging to the different function.

Hypothesis 3: There are N functions in the electronic equipment. There is not coupling between every function, and each function does not share the components. That is: all components in equipment are private components.

The global diagnosis strategy algorithm as follows: if the probabilistic influence for components failure is not considered, and there is not priority for FDI, then any Local diagnosis strategy can be adopted. Form which we can make the root node of the binary tree of local diagnosis strategy as the root node of the global diagnosis strategy algorithm; when all tests in this local D-matrix pass, the root node of the binary tree of the other local diagnosis strategy can be seen as the next node of left branching leaf node of the binary tree of global diagnosis strategy; after all fault may be isolated, the root node of the binary tree of the other local diagnosis strategy can be seen as the next node of right branching leaf node of the binary tree of global diagnosis strategy; Repeating above-mentioned process, until the global diagnosis strategy is formed.

4. Application

Certain equipment is analyzed by the multi-signal model using the TEAMS software [6]. Equipment has functions such as current supply, information display and information processing. From which the functions of current supply and information display are used to illustrate the validity of the method.

The modeling process for the equipment can be got from the Wang[1]. Modeling analysis for the equipment is to get the testability analysis and assessment of the BIT capability. For locating fault more precisely, more test means and test equipment should be needed.

After analyzing and assessing the modeling, the fault diagnosis strategy was generated. When not taking into account local diagnosis strategy, fault diagnosis strategy of the equipment is generated directly. Fault diagnosis strategy of power supply is discontinuous nodes in the global strategy, it is not instructional meaning for maintaining the power supply.

After carrying out structure and the function partition on equipment, adopting local diagnosis strategy, (the fault diagnosis strategy is as follows,

Fig.3.) It is a whole diagnosis strategy.(Fig.3(b). and Fig.3(c). is respectively the right branch and the left branch for the Fig.3(a). in proper order.)For maintenance personal, it is convenient to locate the fault.

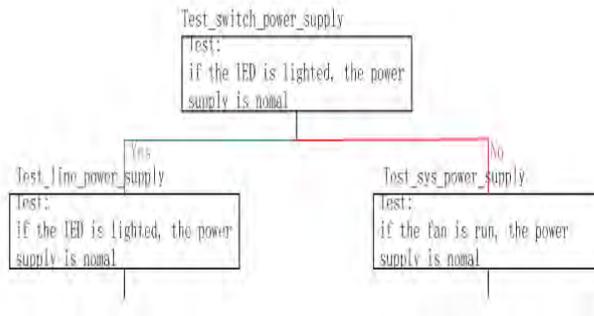


Fig. 3(a).The root node of power supply diagnosis strategy

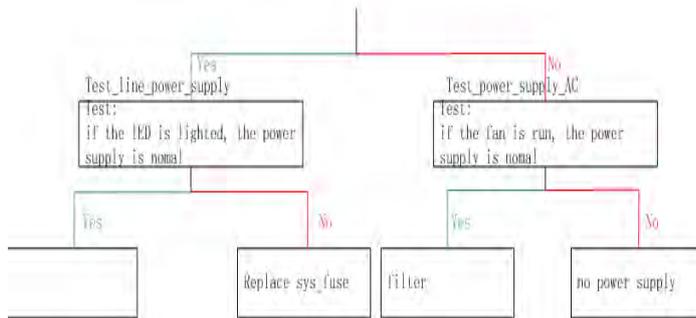


Fig. 3(b). The right branch of power supply diagnosis strategy (Fig. 3(a).the right successive node)

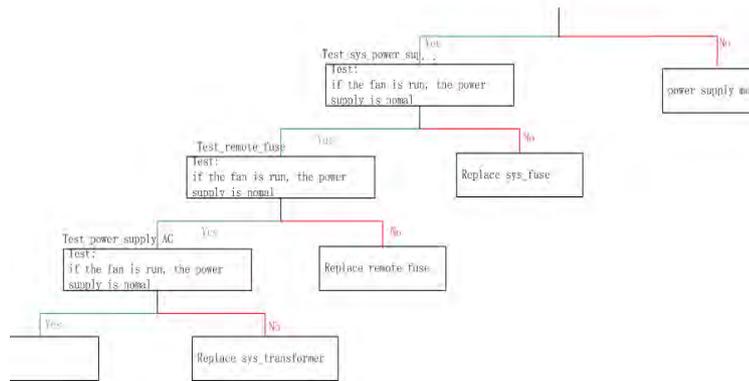


Fig. 3(c). The left branch of power supply diagnosis strategy (Fig. 3(a).the left successive node)

Fig. 3. Power supply diagnosis strategy with local diagnosis strategy

Assume that a complicated electronic equipment has M independent mutually functions, each of which is composed of N different components, furthermore, each function doesn't share the common component. For this system, using the multi-signal model and sorting the test according to information entropy, faults can not be detected and isolated, if the adjacent test is test different functions, the faults can not be detected and isolated. If the diagnosis strategy got from this paper is adopted, the test independent with the function is not called. The computational and storage complexity can be reduced by the optimal and near-optimal test sequencing algorithms for the fault diagnosis strategy.

5. Conclusion

For the distributing of electronic equipment, introducing local diagnosis strategy and adopting appropriate search algorithm, so the global diagnosis strategy is formed. The validity of the method is illustrated by an example from which the conclusion can be drawn that it is an optimal diagnosis strategy which can prove the diagnosticability and maintainability, meanwhile the complexity of computation can be reduced. But if the equipment contain common component in its structure, the extension for the search space should be considered in the following study.

WANG Hong-xia, YE Xiao-hui, and WANG Liang

6. References

1. Wang, H. X., Ye, X. H., Tian, S. X.: Modeling of Fault Diagnosis for Complex Electronic Equipments. Journal of Wuhan University of Technology(Information & Management Engineering),Vol. 29, No. 6, 62-64(2007)
2. Tian, Z.: Determining the Best Diagnostic Procedure by Optimizing the Test Points. Measurement & Control Technology, Vol. 14, No. 4, 12-14 (1995)
3. Jing, X. N., Li, Q. T.: Information Entropy-Based Fault Diagnosis Strategy with Least Test Cost. Computer Applications,Vol. 25, No. 2, 417-419 (2005)
4. Yang, Z. Y., Xu, H. L., Xu, A. Q.: Design of Diagnosis Strategy Based on Multi-signal Model. Computer Measurement & Control,Vol. 14, No. 12, 1616-1619(2006)
5. Qin, Y. A., Jiang, W. J.: Modeling and Application of Complex Diagnosis Distributed Intelligence Based on MAS. Computer Engineering and Applications, No. 29, 142-145 (2004)
6. Dab,S.: Multi-signal Flow Graphs: A novel Approach for System Testability Analysis and Fault Diagnosis. IEEE AUTOTESTCON, 361-373(1994)
7. Mojdeh, S., Vijay, R.: Sequential Testing Algorithms for Multiple Fault Diagnosis . IEEE Transactions on Systems, Man, and Cybernetics---Part A: Systems and Humans,Vol. 1, No. 30, 1-14 (2000)

Wang Hong-xia received B.S degree from Wuhan University in 2003, M.S degree from Naval University of Engineering in 2003 and now she is doctoral candidate in Naval University of Engineering. She is majoring in system testability analysis and fault diagnosis.

Ye Xiao-hui received B.S degree from Naval University of Engineering in 1983, M.S degree and PH.D degree from Naval University of Engineering in 1994 and 2002 respectively, and now he is professor in Naval University of Engineering. he is current research interests fault diagnosis, virtual instrument and system testability analysis.

WANG liang received B.S degree from College of Chongqing Communications of People's Liberation Army in 2005, He is majoring in the TPS and testability analysis.

Received: April 17, 2009; Accepted: September 16, 2009.

3D Point Pattern Matching Based on Spatial Geometric Flexibility

Xiaopeng Wei^{1,2}, Xiaoyong Fang^{1,2}, Qiang Zhang²,
and Dongsheng Zhou^{1,2}

¹ School of Mechanical and Engineering, Dalian University of Technology, Dalian,
116024 China

² Key Laboratory of Advanced Design and Intelligent Computing(Dalian University),
Ministry of Education, Dalian, 116622, China
zhangq26@126.com

Abstract. We propose a new method for matching two 3D point sets of identical cardinality with global similarity but local non-rigid deformations and distribution errors. This problem arises from marker based optical motion capture (Mocap) systems for facial Mocap data. To establish one-to-one identifications, we introduce a forward 3D point pattern matching (PPM) method based on spatial geometric flexibility, which considers a non-rigid deformation between the two point-sets. First, a model normalization algorithm based on simple rules is presented to normalize the two point-sets into a fixed space. Second, a facial topological structure model is constructed, which is used to preserve spatial information for each FP. Finally, we introduce a Local Deformation Matrix (LDM) to rectify local searching vector to meet the local deformation. Experimental results confirm that this method is applicable for robust 3D point pattern matching of sparse point sets with underlying non-rigid deformation and similar distribution.

Keywords. Point pattern matching (PPM), Face model, Spatial geometric flexibility, Topological structure, Motion capture (Mocap), Non-rigid deformation.

1. Introduction

Marker based optical Mocap system is widely used in clinical gait analysis, sports studies, animation and computer games [1], [2], [3]. Passive reflective markers are captured as 3D coordinates by means of image processing. The captured 3D coordinates is unorganized, so there is no method to know which coordinates is correspondent to which marker [4], [5].

This problem could appear in the two steps of Mocap data process: first, model generation, then, noise reduction and missing marker recovery. The second issue of the process for each trail is beyond the scope of our method. Model generation is to build a prior model to track the non-missing data for each sequence. The model is always built base on the first frame data. The

work could be divided into two steps: first, outliers cleaning and missing marker recovery, then, identifications of each marker. Traditionally, this work is all manual[3], [5], [6]. We propose the work to identify the markers to the system model in an automatic process.

This study is focus on the model generation issue in marker-based optical Mocap system, we are aim to present an automatic way to identify the point-set of the first frame model for different subjects or sequences. We formulate this problem as 3D point pattern matching: assume we have a manually identified template point-set $\mathbb{Q}^S = \{\mathbf{Pt}_i^S \in \mathbb{R}^3 | 1 \leq i \leq N\}$, where the symbol S represents the system template model (S-Model), N is the number of the points. We are going to match the to-be-matched point-set to the S-model point-set (M-model) $\mathbb{Q}^M = \{\mathbf{Pt}_i^M \in \mathbb{R}^3 | 1 \leq i \leq N\}$. The to-be-matched model is the first frame model as a prior model for different sequences, the superscript S and M represent the corresponding model. Assume \mathbb{Q}^S and \mathbb{Q}^M have overall spatial distribution similarity and same cardinality. Typically, due to underlying local flexibility and deformation of expressional facial Mocap FPs, there exists no single global scale, nor an affine transformation for the model generation. Considerations should be given to locally non-rigid deformation and special nature of facial structure.

PPM is commonly encountered in computer vision, image analysis, computational geometry, and pattern recognition. While matching subjects existing underlying non-rigid and flexible deformation (like face), non-rigid mapping is called. Besl and McKay [7] present a heuristic *Iterative Closest Point (ICP)* algorithm, which use nearest-neighbor methods to assign correspondence. Rohr et al. [8] and Wahba [9] adopted thin-plate spline to parameterize the non-rigid mapping, in which the outliers do not disturb. The softassign and deterministic annealing [10], [11] formulate pattern matching problem as a heuristic fuzzy problem, which guarantee very trusty one-to-one correspondences. Chui and Rangarajan propose a thin-plate spline based algorithm (TPS-RPM), the TPS-RPM identifies the correspondences, rejects a fraction of the outlier points simultaneously [12], [13]. To fully use the local spatial information, Feng et al. [14] defined the Neighborhood Relative Angle-Context Distribution (NRACD) and used it to match models with underlying local deformation. Zheng and Doermann introduced a point matching algorithm for non-rigid shapes [15], [16]. They formulated local neighborhood structure as a graph, and then used relaxation labeling to refine matching results. In the methods mentioned above, we can see that, the previous work on non-rigid models mostly formulated as an global optimization way[7], [8], [9], [10], [11], [12], [13], while Ref. [14], [15], [16] start to consider the local structure for non-rigid subjects, however, none of them have take local deformation into account for non-rigid and flexible subjects. Considering local deformation, we found that it's a fast and robust way to identify correspondences between 3D point-sets with underlying locally non-rigid deformation and distribution error.

Based above, a spatial geometric flexibility based method is proposed, which considers local deformations between the S-model and M-model. Fig 1. is the flow chart of the algorithm, the LDM is Local Deformation Matrix, which is to correct the local match.

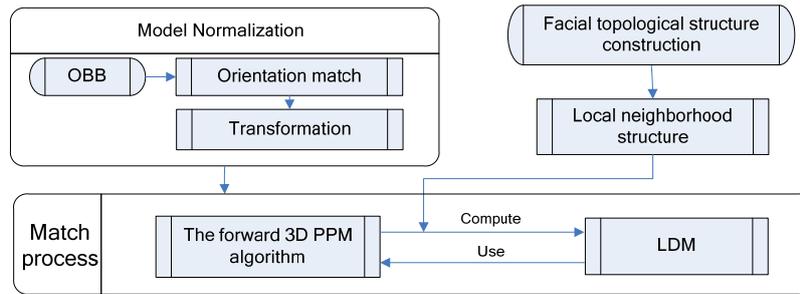


Fig. 1. The flow chart of the algorithm.

2. Model Normalization

Human face is a non-rigid and flexible organic structure, which is comprised of complicated facial muscles and skin organism. To consider local non-rigid deformation, there should be a spatial platform, as the orientation and size are hard to be identical between different FP-sets. So, we proposed normalization method to put the two models into a fixed space.

In traditional way, to match a point with the known correspondence of its neighbor, searching vector is adopted to find the matching correspondence. As the position, size and orientation of the two models are not identical, the searching vectors in matching model (\mathbf{V}_{search}^M) have to transform to register \mathbf{V}_{search}^S , and the registration would repeat for one FP. However, our model normalization method puts the two models into one fixed space, the position, size and orientation of the models are normalized. It simplifies the registrations of searching vectors as all searching vectors are registered in one time, and makes the matching process more clear with low computational cost.

As mentioned above, the normalization could be divided into size and orientation normalization, accordingly, scale and rotation computation. To point-sets with known correspondence, the rotation computation could be formulated as well-known absolute orientation problem [17], [18]. However, in our application, the correspondences of the FP-sets are unknown. So the fundamental issue of the normalization is orientation normalization for the unorganized FP-sets.

2.1. Orientation matching problem

To determine a fixed space for FP-sets, we adopt bounding box technology. The bounding box technology is to find a fittest space for unorganized points.

2.1.1 Bounding box algorithm

The most commonly used bounding box algorithms are AABB (axis-aligned bounding box) [19], [20], Sphere [21], OBB (oriented bounding box) [22], [23] and FDH (fixed directions hulls) [24]. Both AABB and Sphere have not convincing spatial compactness. FDH is derived from AABB which could achieve convincing spatial compactness only by the premise that excessive fixed directions are selected. OBB use an optimal cuboid to bound 3D model, and could possess fine spatial compactness with only 3 orthogonal directions. As a most important fact, the facial topological structure

We choose OBB to construct bounding box for facial FP-sets for the following reasons.

- Fixed space for unorganized points.
- The directions is only 3.
- Geometric vivacity for facial structure: typically, the area of facial front-back side is largest, and the top-down side is smallest (as shown in Fig 2.).

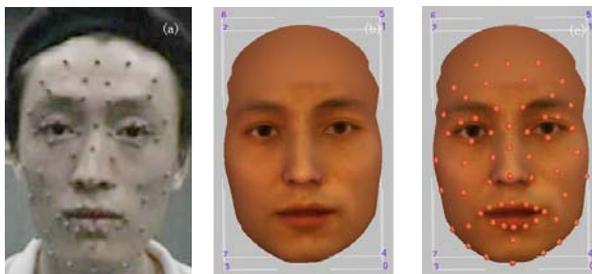


Fig. 2. The real and virtual effect of facial markers' distribution definition.(a) distribution of facial markers; (b)OBB on virtual face model without markers; (c) OBB on virtual face model with markers.

2.1.2 Orientation matching algorithm based on OBB

According to eigenvalue of the covariance matrix constructed by OBB [22], order the corresponding three eigenvectors as \mathbf{m}_x , \mathbf{m}_y , \mathbf{m}_z . \mathbf{m}_x corresponds to the largest eigenvalue. Indicates the eight corner points of the

OBB box as $\mathbf{P}_0, \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4, \mathbf{P}_5, \mathbf{P}_6, \mathbf{P}_7$. \mathbf{P}_O is the center point of the box,

$$\mathbf{P}_O = 0.125 \times \sum_{i=1}^8 \mathbf{P}_i .$$

We use the OBB technology [22] to construct a bounding box for S-Model and M-model. Then it's easy to scale the two boxes into a fixed space. To obtain scale matrix Mat_{Scale} , first, get projection distances $d_{projx}, d_{projy}, d_{projz}$, where $d_{projx} = (\mathbf{P}_6 - \mathbf{P}_0) \odot \mathbf{m}_x$, $d_{projy} = (\mathbf{P}_6 - \mathbf{P}_0) \odot \mathbf{m}_y$, $d_{projz} = (\mathbf{P}_6 - \mathbf{P}_0) \odot \mathbf{m}_z$, the \odot is a vector dot product operator. Initialize Mat_{Scale} to a three order zero matrix, $\mathbf{Mat}_S(1,1) = d_{projx}^S / d_{projx}^M$, $\mathbf{Mat}_S(2,2) = d_{projy}^S / d_{projy}^M$, $\mathbf{Mat}_S(3,3) = d_{projz}^S / d_{projz}^M$. The scale matrix is stored for final normalization.

Fig 3. shows the optimal cuboid and the three directions colored by red (\mathbf{m}_x), green (\mathbf{m}_y) and blue (\mathbf{m}_z) from OBB technology. As the OBB is a statistical methods derived from PCA, it's a black box method. Hence for point-sets under same distribution, each of the three directions may be reversed as long as there are small geometric changes. As shown in Fig 3a. and Fig 3b., the M_x directions are opposite. We call this phenomenon as **orientation inconstancy (OI)**.

Based on OBB technology, we could normalize the orientation of the point-sets by **simple rules**. Firstly, we can first easily determine the nasal tip FP, so we can find a direction that human face towards to; then, to find the second direction, we can find the two points at the lower jaw and near the ear bottom, which is different geometrically from the points on the forehead. We formulate the two rules into algorithm (discuss later), and then the orientation could be normalized.

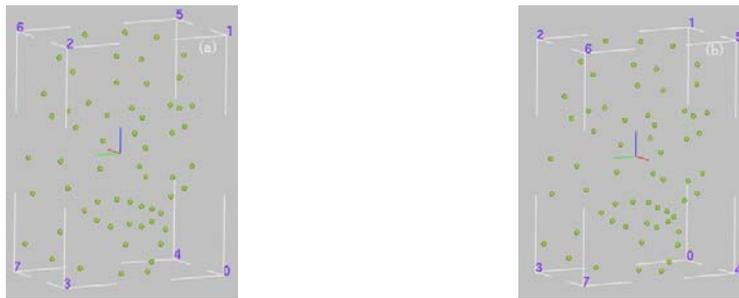


Fig. 3. Orientation inconstancy of First frame data templates. (a) the face facing against the direction \mathbf{m}_x ; (a) the face facing along the direction \mathbf{m}_x .

After the OI had been resolved, the orientation normalization could be formulated as a common rotation computation. It's a common sense in computer graphics, provided that the 3D model's orientation has been given; it does easily obtain a rotation matrix to match orientation of one model to another. Therefore, the orientation matching algorithm is to mainly overcome the OI.

Matching \mathbf{m}_x direction (rule 1)

Step 1: Find the farthest FP along $-M_x$ and M_x in the OBB box of S-model and M-model: P_{xi}^S and P_{xi}^M ($i = 0, 1$).

$$\mathbf{Pt}_{x0} = \mathbf{Pt}_k \Big|_{(1 \leq k \leq N)} (\mathbf{Pt}_k - \mathbf{P}_O) \odot (-\mathbf{m}_x) = \max_{0 \leq j \leq N} (\mathbf{Pt}_j - \mathbf{P}_O) \odot (-\mathbf{m}_x) \quad (1)$$

$$\mathbf{Pt}_{x1} = \mathbf{Pt}_k \Big|_{(1 \leq k \leq N)} (\mathbf{Pt}_k - \mathbf{P}_O) \odot \mathbf{m}_x = \max_{0 \leq j \leq N} (\mathbf{Pt}_j - \mathbf{P}_O) \odot \mathbf{m}_x \quad (2)$$

Step 2: Compute the variance of the distances from \mathbf{Pt}_{xi} ($i = 0, 1$) to from \mathbf{Pt}_{xi} to four OBB vertices $P_{(i*4+j)}$ ($0 \leq j < 4$): var_{xi} ($i = 0, 1$), obey:

$$\bar{d}_{xi} = \frac{1}{4} \sum_{j=0}^3 \|\mathbf{Pt}_{xi} - \mathbf{P}_{4xi+j}\| \quad (3)$$

$$var_{xi} = \frac{1}{3} \sum_{j=0}^3 (\|\mathbf{Pt}_{xi} - \mathbf{P}_{4xi+j}\| - \bar{d}_{xi})^2 \quad (4)$$

Where, \bar{d}_{xi} is the mean value of distances from P_{xi} to the four OBB vertices located in corresponding side of OBB box.

Step 3: Match \mathbf{m}_x direction by comparing var_{x0} to var_{x1} .

Compare var_{x0} to var_{x1} , if the later is smaller, means the FP \mathbf{Pt}_{x1} is more likely nasal tip point, otherwise, $\mathbf{m}_x = -\mathbf{m}_x$, and: $\mathbf{P}_{temp} = \mathbf{P}_j$, $\mathbf{P}_j = \mathbf{P}_{4+j}$, $\mathbf{P}_{4+j} = \mathbf{P}_{temp}$ ($0 \leq j < 4$), where \mathbf{P}_{temp} is a temporary point.

Matching \mathbf{m}_z direction (rule 2)

Step 1: Obtain the nearest FP \mathbf{P}_{zj} ($0 \leq j < 4$) of \mathbf{P}_j ($0 \leq j < 4$):

$$\mathbf{Pt}_{zj} = \mathbf{Pt}_k \Big|_{(1 \leq k \leq N)} (\mathbf{Pt}_k - \mathbf{P}_j) = \min_{0 \leq l \leq N} \{(\mathbf{Pt}_l - \mathbf{P}_j)\} \quad (5)$$

and the $\{\mathbf{P}_j(0 \leq j < 4)\}$ are at the farthest plane along the $-\mathbf{m}_x$ of the OBB box, and the \mathbf{m}_x directions have been normalize. Indicate $d_j^S = \|\mathbf{P}_{zj}^S \mathbf{P}_O^S\|$ ($0 \leq j < 4$) is the distances from \mathbf{P}_{zj} to the center point (\mathbf{P}_O). $\mathbf{V} \in R^{3 \times 1}$, $\mathbf{V}_j^M = \mathbf{P}_{zj}^M - \mathbf{P}_O^M$, $\mathbf{Mat}_{T0}^M = (\mathbf{V}_0^M, \mathbf{V}_1^M, \mathbf{V}_2^M, \mathbf{V}_3^M)$, \mathbf{Mat}_{T1}^M

$= (\mathbf{V}_1^M, \mathbf{V}_0^M, \mathbf{V}_3^M, \mathbf{V}_2^M)$, $\mathbf{Mat}_{Vi}^M = \mathbf{Mat}_{Ti}^M \times \mathbf{Mat}_s (i = 0,1)$. Store the lengths: $d_{ij}^M = \left(\sum_{k=1}^3 (\mathbf{Mat}_{Vi}^M(k, j))^2 \right)^{\frac{1}{2}}$.

$$\bar{d}_{zi} = \frac{1}{4} \sum_{j=0}^3 d_j^S - d_{ij}^M \quad (6)$$

$$var_{zi} = \frac{1}{3} \sum_{j=0}^3 (d_j^S - d_{ij}^M - \bar{d}_{zi})^2 \quad (7)$$

Step 2: Match M_z direction by comparing the two variances.

Compute var_{zi} with formula 3 and 4. Compare var_{z0} and var_{z1} , if the former is smaller, means the states that whether \mathbf{m}_z directs at lower jaw in S-model and M-model are same, otherwise, we force them to be same: $\mathbf{m}_z^M = -\mathbf{m}_z^M$.

After matched \mathbf{m}_x and \mathbf{m}_z , \mathbf{m}_y is easy to match with orthogonality of the three directions: $\mathbf{m}_y = \mathbf{m}_x \otimes \mathbf{m}_z$, where \otimes is a cross-product operator. This orientation matching algorithm is a forward calculation method, with extremely low computational cost, and highly geometric vivacity.

2.2. Transformation

As scale and orientation are given, it's convenient to transform M-model to the S-model space (as shown in Fig.4).



Fig. 4. Demonstration before and after model normalization. (a) before normalization; (b) after normalization

Step 1: Define a matrix \mathbf{Mat}_{Pt}^M for M-model to preserve the FPs' coordinates, wherein, $\mathbf{Mat}_{Pt}^M(i,1) = \mathbf{Pt}_i^M.x$, $\mathbf{Mat}_{Pt}^M(i,2) = \mathbf{Pt}_i^M.y$, $\mathbf{Mat}_{Pt}^M(i,3) = \mathbf{Pt}_i^M.z$ ($1 \leq i \leq N$).

Step 2: Translate M-model to the global coordinate system.

$$\begin{aligned} \mathbf{Mat}_{Pt}^M(i,1) &= \mathbf{Mat}_{Pt}^M(i,1) - \mathbf{P}_O^M \cdot x, & \mathbf{Mat}_{Pt}^M(i,2) &= \mathbf{Mat}_{Pt}^M(i,2) - \mathbf{P}_O^M \cdot y, \\ \mathbf{Mat}_{Pt}^M(i,3) &= \mathbf{Mat}_{Pt}^M(i,3) - \mathbf{P}_O^M \cdot z \end{aligned}$$

Step 3: Rotate M-model to global coordinate system, then scale to uniform the sizes of M-model and S-model. Followed that, rotate M-model to keep its orientation identical to S-model. $\mathbf{Mat}_{RotO}^M = (\mathbf{m}_x^M, \mathbf{m}_y^M, \mathbf{m}_z^M)^{-1}$,

$$\mathbf{Mat}_{RotS}^M = (\mathbf{m}_x^S, \mathbf{m}_y^S, \mathbf{m}_z^S). \mathbf{Mat}_{Pt}^M = \mathbf{Mat}_{RotS}^M \times \mathbf{Mat}_S^M \times \mathbf{Mat}_{RotO}^M \times \mathbf{Mat}_{Pt}^M.$$

Step 4: Translate M-model to the space of S-model.

$$\begin{aligned} \mathbf{Mat}_{Pt}^M(i,1) &= \mathbf{Mat}_{Pt}^M(i,1) + \mathbf{P}_O^S \cdot x, & \mathbf{Mat}_{Pt}^M(i,2) &= \mathbf{Mat}_{Pt}^M(i,2) + \mathbf{P}_O^S \cdot y, \\ \mathbf{Mat}_{Pt}^M(i,3) &= \mathbf{Mat}_{Pt}^M(i,3) + \mathbf{P}_O^S \cdot z. \end{aligned}$$

Back-feed the coordinate matrix to the FPs' coordinates, $\mathbf{Pt}_i^M \cdot x = \mathbf{Mat}_{Pt}^M(i,1)$, $\mathbf{Pt}_i^M \cdot y = \mathbf{Mat}_{Pt}^M(i,2)$, $\mathbf{Pt}_i^M \cdot z = \mathbf{Mat}_{Pt}^M(i,3)$ ($1 \leq i \leq N$), as shown in Fig. 4b.

After normalization, the influence of orientation and size difference between S-model and M-model is removed, so it is independent of orientation and size.

3. Construction of human facial topological structure

The facial organs such as mouth, eyes are unclosed, the movements of the upper and lower mouth lips are always different largely (same as the eye lips). Therefore, it's unreasonable to consider human face as a whole. Based on this consideration, firstly, we use Delaunay triangulation technology to construct a facial surface, and then neighborhood structures for FPs are obtained, as shown in Fig 5a. Secondly, by human computer interaction technology, we remove improper neighborhood relation, as shown in Fig. 5b. and 5c. Fig. 5a. shows some improper neighborhood relations between upper lower eyelids, upper lower lips, nasal tip and upper lip. Only with the structure considered the openness of eyes and mouth, the local deformation could be computed in nature.

3.1. 3D Triangulation and improper relationship removal

We use a simple 3D triangulation to construct the surface. First, we project all the FPs to the front plane of the OBB box, whose normal is \mathbf{m}_x . Then the Delaunay triangulation is processed in 2D. To remove improper relationship, the human machine technology is used. We use an OOP (object oriented programming) in VC++, all the FP and relation lines are a pickable object. We developed a tool to remove or add relation lines for the facial surface. The particulars are beyond the scope of the paper.

3.2. Obtain local neighborhood relations

With the constructed facial topological structure, the neighbors of each FP are known. A relation line is defined to connect two adjacent points in the topological structure. Traverse the relation lines to obtain neighboring FPs' indices for each facial FP. The local neighborhood structure includes spatial relation such as angles, distances, orientations and so on.

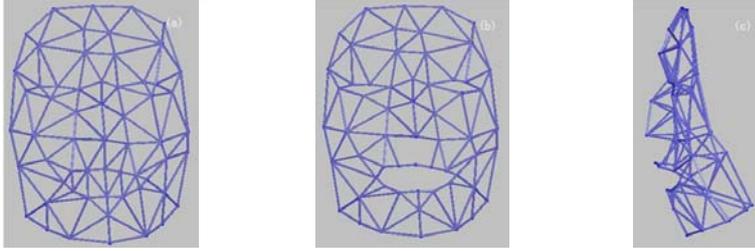


Fig. 5. Facial topological structure definition. (a) before removing improper relations after triangulation ;(b) facial topological structure, front view; (c) facial topological structure, side view.

4. Robust 3D point pattern matching based on spatial geometric flexibility

Def 1: let current base point index is $i_{curbase}^S$, current matching point index be $i_{curmatch}^S$, the searching vector in S-model be $\mathbf{V}_{Search}^S = \mathbf{Pt}_{i_{curmatch}^S}^S - \mathbf{Pt}_{i_{curbase}^S}^S$, LDM be \mathbf{Mat}_{LDM} , and searching radius be $r_{i_{curmatch}^S}^S$, if there exists a point whose index is $i_{curmatch}^M$ and it is in the sphere with center $\mathbf{Pt}_{i_{curmatch}^M}^M$ ($\mathbf{Pt}_{i_{curmatch}^M}^M = \mathbf{Pt}_{i_{curbase}^M}^M + \mathbf{Mat}_{LDM} \times \mathbf{V}_{Search}^S$) and radius $r_{i_{curmatch}^S}^S$, then $i_{curbase}^S$ and $i_{curmatch}^M$ constitute a point match $\begin{pmatrix} i_{curmatch}^S \\ i_{curmatch}^M \end{pmatrix}$, we call this temporary success match (TSM) of $\mathbf{Pt}_{i_{curmatch}^S}^S$.

Def 2: let the size of the indices vector $vec_{i_{curbase}^S}^S$ (preserve the local neighborhood points' indices of $i_{curbase}^S$) be $N_{veccurbase}^S$, and the TSM number of points correspond to $vec_{i_{curbase}^S}^S$ be N_{TSM}^S , if $N_{TSM}^S / N_{veccurbase}^S > \varepsilon_{submatch}$, we call

the point match $\begin{pmatrix} i_{curbase}^S \\ i_{curbase}^M \end{pmatrix}$ final success match (FSM) of $\mathbf{Pt}_{i_{curbase}}^S$. We call the

way determining FSM by the relative amount of TSMs Temporary Feedback Method (TFM).

After normalization, the flow of robust 3D point pattern match algorithm based on spatial geometric flexibility is programmed in following pseudocode:

```

Push nasal tip FP's index into StackTempBase ( $i_{nose}^S$  to TempBaseStackS and  $i_{nose}^M$  to TempBaseStackM), Loop Time =0;
While TempBaseStack ≠ Null
  For i = 1 to size of StackTempBase
     $i_{curbase}$  = TempBaseStack[i];
    Compute and preserve TSMs for current base point
     $\mathbf{Pt}_{i_{curbase}}^S$ ;
    Load information of local topological structure
    If Loop Time ≠ 0
      If  $N_{TSM}^S / N_{veccurbase}^S > \zeta_{submatch}$ 
         $\begin{pmatrix} i_{curmatch}^S \\ i_{curmatch}^M \end{pmatrix}$  is confirmed as a FSM.
      Endif
    Endif
    Update LDM for base point, and distribute the LDM;
    Push  $i_{curbase}^S$  into TempStack;
  EndFor
  Displace TempBaseStack by TempStack, empty TempStack,
  plus 1 to Loop Time.
EndWhile
Empty TempBaseStack.

```

To insure the robustness of each point correspondence, we use a double insurance method. Firstly, we use known correspondence to search neighboring FP \mathbf{Pt}_j^M corresponding to \mathbf{Pt}_i^M ($i, j < M$). Secondly, we use Temporary Feedback Method to confirm a final match (discuss later). Above all, we assign an identity matrix to every LDM, which actually can not

represent local deformation. The first FSM $\begin{pmatrix} i_{nose}^S \\ i_{nose}^M \end{pmatrix}$, is to be used as reference

point correspondence. As the FSM is increasing, the LDM is updated to meet the local deformation. If the FSMs of the current base point is no less than two, we can compute the LDM, then the local deformation could be used for matching.

4.1. Algorithm of LDM computation

After normalization, the size and orientation between S-model and M-model are identical. As the FSM increasing in matching process, known correspondences are increasing. With these known correspondences, we formulate local rotation as a least square optimization:

$$\operatorname{argmin}_{\mathbf{R}_i} \sum_{k \in Y_i} \|\mathbf{R}_i \cdot (\mathbf{Pt}_{i,k}^M - \mathbf{Pt}_i^M) - (\mathbf{Pt}_{i,k}^S - \mathbf{Pt}_i^S)\|^2 \quad (8)$$

where, we denote the indices of one-ring neighbor for marker i as Y_i , $\mathbf{Pt}_{i,k}$ is the k th (in Y_i) neighbor of the i th marker. $\mathbf{R}_i^t \in \mathfrak{R}^{3 \times 3}$ indicates the optimum rotation for marker i respectively. We adopt the method of Horn [25] to find the optimum rotation matrix \mathbf{Mat}_{LDM} for i th marker: \mathbf{R}_i . Besides the rotation, the space for tensile deformation is committed to the dynamic searching radius. However, Horn's method could not solve the situation the neighborhood points are less than three [18]. The minimum number of neighborhood points for every FP is two. So, when neighborhood points for current base point is two, just set $\varepsilon_{submatch}$ bigger than 0.5, we can insure the current base point is FSM (because it requires more than 0.5×2 neighborhood points are TSM. In our experiments, we set it to be 0.67). Based above, a simple approach is proposed to compute LDM with the two neighboring FPs.

Step 1: Construct a local plane with current base point and the two neighboring FPs. Obtain normal \mathbf{V}_{norm} of the plane. Let the closest point of

$\mathbf{Pt}_{i_{curbase}}$'s index be i_{near} , and the other's index be i_{far} , $\mathbf{V}_{near} = \mathbf{Pt}_{i_{near}} -$

$\mathbf{Pt}_{i_{curbase}}$, $\mathbf{V}_{far} = \mathbf{Pt}_{i_{far}} - \mathbf{Pt}_{i_{curbase}}$, so:

$$\mathbf{V}_{norm} = \mathbf{V}_{near} \times \mathbf{V}_{far} \quad (9)$$

Step 2: Compute the orthogonal vector of \mathbf{V}_{near} and $\mathbf{V}_{norm} : \mathbf{V}_{cons}$, then compute the LDM:

$$\mathbf{V}_{cons} = \mathbf{V}_{near} \times \mathbf{V}_{norm} \quad (10)$$

$$\mathbf{Mat}_{LDM} = (\mathbf{V}_{cons}^M, \mathbf{V}_{near}^M, \mathbf{V}_{norm}^M) \times (\mathbf{V}_{cons}^S, \mathbf{V}_{near}^S, \mathbf{V}_{norm}^S)^{-1} \quad (11)$$

4.2. Local searching with dynamic searching radius

Use the local topological structure we preserved to get the closet point \mathbf{Pt}_j^S of \mathbf{Pt}_i^S ($i \neq j$), $D_i = \|\mathbf{Pt}_j^S - \mathbf{Pt}_i^S\|$. We set a searching threshold value ε_r ,

($0 < \varepsilon_r < 1$) to control searching range for each FP: $r_i = \varepsilon_r \cdot D_i$. Basically, the bigger ε_r is, the better tolerance to data distribution error our method has. According to experience from our experiments, ε_r is recommended to be close to 0.5.

5. Experiments and results

In our experiments, we use 60 markers for facial marker setup (as shown in Fig 2.). To evaluate the performance of our algorithm, we select five M-models which are with expressions: narrowing one eye, surprise, eyebrow up, laughing exaggeratedly, mouth and eyes widely open respectively (as shown in Fig 6b. to Fig 6f.), and we select a model with neutral expression (as shown in Fig 6a.) as our S-model. The algorithm described in the preceding sections has been implemented in a Visual C++ program. Our experiments were run on a PC with two 1.86GHz Pentium IV processors and 2GB memory.

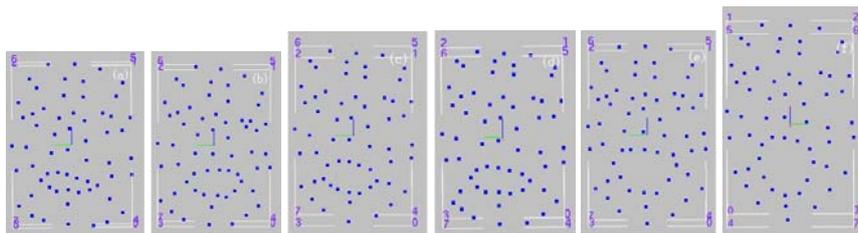


Fig. 6. First frame data templates with different expressions. (a) neutral;(b)narrowing one eye;(c)surprise;(d)eyebrow up;(e)laughing exaggeratedly;(f)mouth and eyes widely open.

We first show the effectiveness of our normalization method. Fig 7. shows the visual effects of experiment results, where red colored model is represents the normalized M-model. As we can see from this Figure, no matter how big the expressional difference between S-model and M-model is, the performance of our normalization algorithm is acceptable.

Next, we compare our FP matching algorithm with two different ways: using LDM (Fig 7c., Fig 7e. and Fig 7g.) and without using LDM (Fig 7d., Fig 7f. and Fig 7h.). Table 1. shows the result or comparison, wherein **FSM** column shows the number of final successful matched point, **Time** column shows machine running time. In addition, ε_r column represents the value which starts from zero and satisfies that 100% FPs are FSM. From this table, to models with expression which has slight local motions, such as the model with expression narrowing one eye, we use our algorithm without using LDM. Because local neighborhood structure was considered, it still achieved good results even with less computational cost. Along with increasing of the local

motions' extent, it's easy to break in cycle process (as shown in Fig 7d., Fig 7f. and Fig 7h.). The reason is our method without using LDM didn't consider the effect of local motions and deformations. Due to local motions and deformations are considered, our algorithm with using LDM could match FPs for facial Mocap frame data templates almost perfectly. Although our algorithm with using LDM pays more computational cost, it still solves non-rigid and flexible matching in a considerable time. Some models with special or nearly extreme local deformations such as the model with expression: mouth and eyes widely open (as shown in Fig 6f.), are perfectly matched with S-model in our algorithm with using LDM.

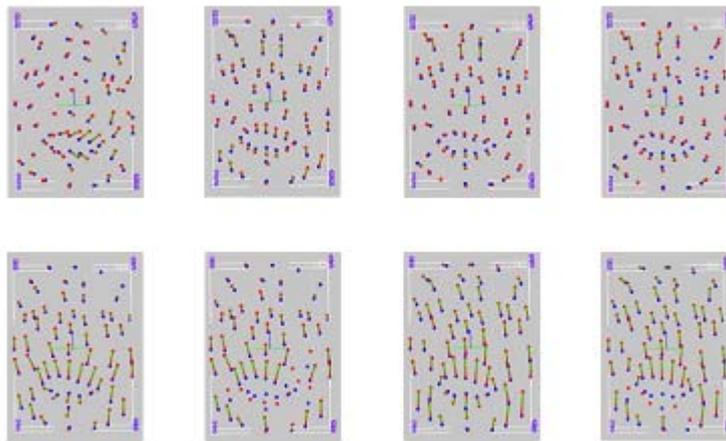


Fig. 7. The matching results of first frame data templates with different expressions.(a) the result of matching M1-model; (b) the result of matching M2-model; (c) the result of matching M3-model using LDM; (d) the result of matching M3-model without using LDM; (e) the result of matching M4-model using LDM; (f) the result of matching M4-model without using LDM; (g) the result of matching M5-model using LDM; (h) the result of matching M5-model without using LDM.

6. Conclusion

This paper proposed a robust PPM for matching unorganized point-sets with underlying non-rigid deformation. Our method is a forward calculation method with highly geometric vivacity. To consider local deformation, we normalize the point-sets into a fixed space, and then check the minor difference to derive local deformation. With the local deformation, the calculated LDM to rectify searching space, we achieved a more robust matching effort than considering local structure without deformation. This method considers the nature of the distribution of the FPs to find some rules to normalize the orientation. This model normalization is geometric intuitive, but can hardly be

extended to all non-rigid point-sets, only if we can find some rules from the point distribution. The proposed method has been found effective for solving 3D facial point pattern matching problem for marker-based optical Mocap system. Experimental results demonstrate the robustness in different situation, with considerable low amount of computational cost.

Table 1. The experiment results of the Facial point matching algorithm

Expressions in Matching Models (M-model)	Using LDM			Without using LDM			Searching threshold value(ϵ_r)
	FSM	Loop Time	Time (ms)	FSM	Loop Time	Time (ms)	
Narrowing one eye (Fig 76)	60	9	47	60	7	32	0.303
Surprise (Fig 6c)	60	7	47	60	9	32	0.340
Eyebrow up (Fig 6d)	60	9	47	55	7	32	0.192
Laughing exaggeratedly (Fig 6e)	60	9	31	47	8	31	0.312
Mouth and eyes widely open (Fig 6f)	60	8	31	51	8	31	0.348

7. Acknowledgment

This work is supported by National Natural Science Foundation of China (Grant No. 60875046), by Program for New Century Excellent Talents in University (Grant No. NCET-06-0298), by the Key Project of Chinese Ministry of Education.(No.209029), by China Postdoctoral Science Foundation ((Grant No. 2009045 1266), by the Program for Liaoning Innovative Research Team in University (Grant Nos.2008T004 and 2009T005), by the Program for Liaoning Science and Technology Research in University (Grant Nos.2009S008, 2009S009) and by the Program for Liaoning Excellent Talents in University (Grant No.2006R06).

8. References

1. Moeslund, T. B. and Granum, E.: A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*. 81, 231 - 268. (2001)
2. Moeslund, T. B. and Hilton, A. and Krüger, V.: A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*. 104, 90 - 126. (2006)

3. Richards, J. G.: The measurement of human motion: A comparison of commercially available systems. *Human Movement Science*. 18, 589-602. (1999)
4. Gleicher, M.: Animation from Observation: Motion Capture and Motion Editing. *Computer Graphics*. 33, 4, 51-54. (1999)
5. Li, B. and Holstein, H.: Using k-d Trees for Robust 3D Point Pattern Matching. *Proceedings of Fourth International Conference on 3D Digital Imaging and Modeling*, 95–102. (2003)
6. Jobbagy, A., Furnee, E., Harcos, P. Tarczy, M. Early detection of Parkinson's disease through automatic movement evaluation. *Engineering in Medicine and Biology Magazine*. 1998, 17(2):81-88
7. Besl, P. J. and McKay, D.: A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14 ,2, 239-256. (1992)
8. Rohr, K. and Stiehl, H.S. and Sprengel, R. and Beil, W. and Buzug, T.M. and Weese, J. and Kuhn, M.H.: Point-Based Elastic Registration of Medical Image Data Using Approximating Thin-Plate Splines. *Visualization in Biomedical Computing*. 1131, 297-306. (1996)
9. Wahba, G.: *Spline models for observational data*. SIAM, Philadelphia, PA. (1990)
10. Chui, H. and Rambo, J. and Duncan, J. and Schultz, R. and Rangarajan, A.: Registration of cortical anatomical structures via robust 3Dpoint matching. *Information Processing in Medical Imaging (IPMI)*, Springer, Berlin, 168–181. (1999)
11. Gold, S. and Rangarajan, A. and Lu, C.P. and Pappu, S. and Mjolsness, E.: New algorithms for 2-D and 3-D point matching: pose estimation and correspondence. *Pattern Recognition*. 31, 8, 1019–1031. (1998)
12. Chui, H and Rangarajan, A.: A new algorithm for non-rigid point matching. *Computer Vision and Pattern Recognition*. 2, 44-51. (2000)
13. Chui, H and Rangarajan, A.: A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*. 2003,89: 114–141
14. Feng, J. and Horace, H and Ip, S. and Lai, L.Y. and Linney, A.: Robust point correspondence matching and similarity measuring for 3D models by relative angle-context distributions. *Image and Vision Computing*. 26, 6, 761-775. (2008)
15. Zheng, Y. and Doermann, D.: Robust Point Matching for Non-Rigid Shapes: A Relaxation Labeling Based Approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 4, 643-649. (2004)
16. Zheng, Y. and Doermann, D.: Robust Point Matching for Nonrigid Shapes by Preserving Local Neighborhood Structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 28, 4, 643-649. (2006)
17. Kanatani, K.: Analysis of 3-D Rotation Fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 16, 5, 543-549. (1994)
18. Park, S. and Hodgins, J. K.: Capturing and animating skin deformation in human motion, *ACM Transactions on Graphics (TOG)*. July, 25, 3. (2006)
19. Ponamgi, M. and Manocha, D. and Lin, M.: Incremental Algorithms for Collision Detection Between Polygonal Models. *IEEE Transactions on Visualization and Computer Graphics*. 3, 1, 51-64. (1997)
20. van den Bergen, G.: Efficient collision detection of complex deformable models using AABB trees. *Journal of Graphics Tools*. 2, 4, 1-13. (1997)
21. Hubbard, P. M.: Collision Detection for Interactive Graphics Applications. *IEEE Transaction on Visualization and Computer Graphics*. 1, 3, 218-230. (1995)
22. Gottschalk, S. and Lin, M. C. and Manocha, D.: OBB tree: A hierarchical structure for rapid interference detection. *Proceedings of the ACM SIGGRAPH 1996*. New Orleans, USA, 171-180. (1996)

Xiaopeng Wei, Xiaoyong Fang, Qiang Zhang, and Dongsheng Zhou

23. Barequet, G. and Har-Peled, S.: Efficiently approximating the minimum-volume bounding box of a point set in three dimensions. *Journal of Algorithms*. 38, 1, 91-109. (2001)
24. Kay, T. L. and Kajiya, J. T.: Ray Tracing Complex Scenes. *Computer Graphics*. 20, 4, 269-278. (1983)
25. Horn, B. K. P.: Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*. 5, 7, 1127-1135. (1988)

Xiaopeng Wei is a professor at Dalian University of Technology and Dalian University, Dalian, China. His research area include computer animation, intelligent CAD. So far, he has (co-) authored about 160 papers published.

Xiaoyong Fang and **Dongsheng Zhou** are Ph.D students at Dalian University of Technology. Their research direction is computer animation.

Qiang Zhang is a professor at Dalian University, Dalian, China. His research interests are computer animation, intelligent computing. Now he has served as editorial boards of seven international journals and has edited special issues in journals such as *Neurocomputing* and *International Journal of Computer Applications in Technology*.

Received: April 25, 2009; Accepted: October 26, 2009.

CIP – Каталогизacija y publikaciji
Народна библиотека Србије, Београд

004

COMPUTER Science and Information
Systems : the International journal /
Editor-in-Chief **Mirjana Ivanović**. – Vol. 7,
No 1 (2010) - . – Novi Sad (**Trg D. Obradovića**
3): ComSIS Consortium, 2010 - (Belgrade
: Sgra star). –30 cm

Polugodišnje. – Tekst na engleskom jeziku

ISSN 1820-0214 = Computer Science and
Information Systems
COBISS.SR-ID 112261644

Cover design: V. Štavljanin
Library Administration: M. Petković
Printed by: Sgra star, Beograd



Com
SIS

Computer Science
and Information Systems

Volume 7, Number 1, February 2010

Contents

Editorial

Guest Editorial

Papers

- 1 An Extended Photometric Stereo Algorithm for Recovering Specular Object Shape and Its Reflectance Properties
Zuoyong Zheng, Lizhuang Ma, Zhong Li, and Zhihua Chen
- 13 Extracting PCB Components Based on Color Distribution of Highlight Areas
Zhou Zeng, LiZhuang Ma, and ZuoYong Zheng
- 31 Service-oriented Enterprise Interoperability in Automobile Supply Chain Management
Yong Zhang, Shijun Liu, Lei Wu, and Yuchang Jiao
- 51 Modeling and Simulation of a Spherical Mobile Robot
Sang Shengju, Zhao Jichao, Wu Hao, Chen Shoujun, and An Qi
- 63 3D Mesh Skeleton Extraction Using Prominent Segmentation
Xiaopeng Sun, J. Pan, and Xiaopeng Wei
- 75 A Dynamic Alignment Algorithm for Imperfect Speech and Transcript
Ye Tao, Xueqing Li, and Bian Wu
- 85 Multi-Video Summarization Using Complex Graph Clustering and Mining
Jian Shao, Dongming Jiang, Mengru Wang, Hong Chen, and Lu Yao
- 99 A Retrieval Method for Human Mocap Data Based on Biomimetic Pattern Recognition
Xiaopeng Wei, Boxiang Xiao, and Qiang Zhang
- 111 Generative 3D Images in a Visual Evolutionary Computing System
Hong Liu
- 127 Effective Semi-supervised Nonlinear Dimensionality Reduction for Wood Defects Recognition
Zhao Zhang and Ning Ye
- 139 MFI-Tree: An Effective Multi-feature Index Structure for Weighted Query Application
Yunfeng He and Junqing Yu
- 153 A Content-based Dynamic Load-Balancing Algorithm for Heterogeneous Web Server Cluster
Zhang Lin, Li Xiao-ping, and Su Yuan
- 163 Blind Separation Using Second Order Statistics for Non-stationary Signals
Jun Du, Ju Liu, Shengju Sang, and Jun Wang
- 177 An Accelerometer-Based Gesture Recognition Algorithm and its Application for 3D Interaction
Jianfeng Liu, Zhigeng Pan, and Xiangcheng Li
- 189 The Design and Evaluation of Hierarchical Multi-level Parallelisms for H.264 Encoder on Multi-core Architecture
Haitao Wei, Junqing Yu, and Jiang Li
- 201 Robust Moving Object Detection Under Complex Background
Ding Ying, Li Wen-hui, Fan Jing-tao, and Yang Hua-min
- 211 A Novel Hierarchical Speech Emotion Recognition Method Based on Improved DDAGSVM
Qi-rong Mao and Yong-zhao Zhan
- 223 Research on Optimizing the Fault Diagnosis Strategy of Complex Electronic Equipment
Wang Hong-xia, Ye Xiao-hui, and Wang Liang
- 231 3D Point Pattern Matching Based on Spatial Geometric Flexibility
Xiaopeng Wei, Xiaoyong Fang, Qiang Zhang, and Dongsheng Zhou
-

ISSN: 1820-0214