

Tieniu Tan Qiuqi Ruan
Shengjin Wang Huimin Ma
Kaiqi Huang (Eds.)

Communications in Computer and Information Science

437

Advances in Image and Graphics Technologies

Chinese Conference, IGTA 2014
Beijing, China, June 19–20, 2014
Proceedings

Editorial Board

Simone Diniz Junqueira Barbosa

*Pontifical Catholic University of Rio de Janeiro (PUC-Rio),
Rio de Janeiro, Brazil*

Phoebe Chen

La Trobe University, Melbourne, Australia

Alfredo Cuzzocrea

ICAR-CNR and University of Calabria, Italy

Xiaoyong Du

Renmin University of China, Beijing, China

Joaquim Filipe

Polytechnic Institute of Setúbal, Portugal

Orhun Kara

TÜBİTAK BİLGEM and Middle East Technical University, Turkey

Igor Kotenko

*St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences, Russia*

Krishna M. Sivalingam

Indian Institute of Technology Madras, India

Dominik Ślęzak

University of Warsaw and Infobright, Poland

Takashi Washio

Osaka University, Japan

Xiaokang Yang

Shanghai Jiao Tong University, China

Tieniu Tan Qiuqi Ruan Shengjin Wang
Huimin Ma Kaiqi Huang (Eds.)

Advances in Image and Graphics Technologies

Chinese Conference, IGTA 2014
Beijing, China, June 19-20, 2014
Proceedings

 Springer

Volume Editors

Tieniu Tan
Chinese Academy of Sciences, Beijing, China
E-mail: tnt@nlpr.ia.ac.cn

Qiuqi Ruan
Beijing Jiaotong University, China
E-mail: qqruan@center.njtu.edu.cn

Shengjin Wang
Tsinghua University, Beijing, China
E-mail: wsgsj@tsinghua.edu.cn

Huimin Ma
Tsinghua University, Beijing, China
E-mail: mhmpub@mail.tsinghua.edu.cn

Kaiqi Huang
Chinese Academy of Sciences, Beijing, China
E-mail: kqhuang@nlpr.ia.ac.cn

ISSN 1865-0929

ISBN 978-3-662-45497-8

DOI 10.1007/978-3-662-45498-5

Springer Heidelberg New York Dordrecht London

e-ISSN 1865-0937

e-ISBN 978-3-662-45498-5

Library of Congress Control Number: 2014953784

© Springer-Verlag Berlin Heidelberg 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

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

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

It was a pleasure and an honor to have organized the 8th Conference on Image and Graphics Technologies and Applications. The conference was held on June 19–20, 2014, in Beijing, China. The conference series is the premier forum for presenting research in image processing and graphics and their related topics. The conference provides a platform for sharing the progress in these areas: the generation of new ideas, new approaches, new techniques, new applications, and new evaluation. The conference is organized under the auspices of the Beijing Society of Image and Graphics.

The conference program includes keynotes, oral papers, posters, demos and exhibitions. For this year's conference, we received 110 papers for review. Each of these was assessed by no fewer than two reviewers, with some of the papers being assessed by three reviewers; 39 submissions were selected for oral and poster presentation.

We are grateful for the efforts of everyone who helped make this conference a reality. We received a record number of submissions this year and we are grateful to the reviewers, who completed the reviewing process on time. The local host, the Academy of Armored Forces Engineering, enabled many of the local arrangements for the conference.

The conference continues to provide a leading forum for cutting-edge research and case studies in image and graphics.

June 2014

Qiuqi Ruan

Table of Contents

Image Quality Assessment Based on SIFT and SSIM	1
<i>Wenjun Lu, Congli Li, Yongchang Shi, and Xiaoning Sun</i>	
Edge Detection in Presence of Impulse Noise	8
<i>Yuying Shi, Feng Guo, Xinhua Su, and Jing Xu</i>	
A Novel Multi-focus Image Capture and Fusion System for Macro Photography	19
<i>Yili Zhao, Yi Zhou, and Dan Xu</i>	
Comparative Study of Near-Lossless Compression by JPEG XR, JPEG 2000, and H.264 on 4K Video Sequences	29
<i>Wang Dan</i>	
A Retinex-Based Local Tone Mapping Algorithm Using L_0 Smoothing Filter	40
<i>Lei Tan, Xiaolin Liu, and Kaichuang Xue</i>	
A Fast Mode Decision Algorithm and Its Hardware Design for H.264/AVC Intra Prediction	48
<i>Wei Wang, Yuting Xie, Tao Lin, and Jie Hu</i>	
Research of Multi-focus Image Fusion Algorithm Based on Sparse Representation and Orthogonal Matching Pursuit	57
<i>Li Xuejun and Wang Minghui</i>	
Infrared Face Recognition Based on DCT and Partial Least Squares	67
<i>Zhihua Xie and Guodong Liu</i>	
Pipeline Architecture for High Speed License Plate Character Recognition	74
<i>Boyu Gu, Qiang Zhang, and Zhenhuan Zhao</i>	
Robust Dual-Kernel Tracking Using Both Foreground and Background	83
<i>Wangsheng Yu, Zhiqiang Hou, Xiaohua Tian, Lang Zhang, and Wanjuan Xu</i>	
Humanoid-Eye Imaging System Model with Ability of Resolving Power Computing	91
<i>Ma Huimin and Zhou Luyao</i>	
Color Cast Detection Method Based on Multi-feature Extraction	103
<i>Minjing Miao, Yuan Yuan, Juhua Liu, and Hanfei Yi</i>	

Research on an Extracting Method to Assist Manual Identification of Targets	110
<i>Yue Shi and Guoying Zhang</i>	
A Face Recognition under Varying Illumination	120
<i>Haodong Song, Xiaozhu Lin, and Zhanlin Liu</i>	
CS-FREAK: An Improved Binary Descriptor	129
<i>Jianyong Wang, Xuemei Wang, Xiaogang Yang, and Aigang Zhao</i>	
Lung Fields Segmentation Algorithm in Chest Radiography	137
<i>Guodong Zhang, Lin Cong, Liu Wang, and Wei Guo</i>	
Automatic Classification of Human Embryo Microscope Images Based on LBP Feature	145
<i>Liang Xu, Xuefeng Wei, Yabo Yin, Weizhou Wang, Yun Tian, and Mingquan Zhou</i>	
Robust and Accurate Calibration Point Extraction with Multi-scale Chess-Board Feature Detector	153
<i>Yang Liu, Yisong Chen, and Guoping Wang</i>	
A Mixed-Method Approach for Rapid Modeling of the Virtual Scene Building	165
<i>Pu Ren, Wenjian Wang, Mingquan Zhou, and Chongbin Xu</i>	
An Improved Method of Building Rapid 3D Modeling Based on Digital Photogrammetric Technique	175
<i>Zimin Zhang, Ying Zhou, Jian Cui, and Hao Liu</i>	
3-D Reconstruction of Three Views Based on Manifold Study	181
<i>Li Cong, Zhao Hongrui, Fu Gang, and Peng Xingang</i>	
Distribution and Rendering of Real-Time Scene Based on the Virtual Reality System	192
<i>Chengfang Zhang and Guoping Wang</i>	
Probabilistic Model for Virtual Garment Modeling	203
<i>Shan Zeng, Fan Zhou, Ruomei Wang, and Xiaonan Luo</i>	
Dense 3D Reconstruction and Tracking of Dynamic Surface	213
<i>Jinlong Shi, Suqin Bai, Qiang Qian, Linbin Pang, and Zhi Wang</i>	
Scene Simulation for a New Type of Armored Equipment Simulator	224
<i>Guanghui Li, Qiang Liang, Wei Shao, and Xu-dong Fan</i>	
Study on the Key Technique of the Hill Shading Virtual Roaming System Based on XNA	230
<i>Hesong Lu, Zaijiang Tang, Qiang Liang, and Wei Shao</i>	

Study on MMO in Visual Simulation Application	239
<i>Liang Qiang, Fan Rui, Xu Renjie, and Du Jun</i>	
Towards the Representation of Virtual Gymnasia Based On Multi-dimensional Information Integration	249
<i>Xiangzhong Xu, Jiandong Yang, Haohua Xu, and Yaxin Tan</i>	
Night Vision Simulation of Drive Simulator Based on OpenGVS4.5	257
<i>Zheng Changwei, Xue Qing, and Xu Wenchao</i>	
The Design and Implementation of Military Plotting System Based on Speech Recognition Technology	264
<i>Wei Shao, Guanghui Li, Xiyong Huang, Qiang Liang, and Hesong Lu</i>	
Defect Detection in Fabrics Using Local Binary Patterns	274
<i>Pengfei Li, Xuan Lin, Junfeng Jing, and Lei Zhang</i>	
Research on Teeth Positioning Based on Wavelet Transform and Edge Detection	284
<i>Zhou Zhou, Guoxia Sun, and Tao Yang</i>	
The Orientation Angle Detection of Multiple Insulators in Aerial Image	294
<i>Zhenbing Zhao, Ning Liu, Mingxiao Xi, and Yajing Yan</i>	
Improved Robust Watermarking Based on Rational Dither Modulation	305
<i>Zairan Wang, Jing Dong, Wei Wang, and Tieniu Tan</i>	
The Research of Vehicle Tracking Based on Difference Screening and Coordinate Mapping	315
<i>Zhang Jun-yuan, Liu Wei-guo, Tong Bao-feng, and Wang Nan</i>	
Pathology Image Retrieval by Block LBP Based pLSA Model with Low-Rank and Sparse Matrix Decomposition	327
<i>Yushan Zheng, Zhiguo Jiang, Jun Shi, and Yibing Ma</i>	
Remote Sensing Image Change Detection Based on Low-Rank Representation	336
<i>Yan Cheng, Zhiguo Jiang, Jun Shi, Haopeng Zhang, and Gang Meng</i>	
Two New Methods for Locating Outer Edge of Iris	345
<i>Yujie Liu and Hong Tian</i>	
Effects of the Gridding to Numerical Simulation of the Armour-Piercing Warhead Penetration through the Steel Target	352
<i>Jun-qing Huang, Ya-long Fan Rui Ma, and Wei Shao</i>	
Author Index	359

Image Quality Assessment Based on SIFT and SSIM

Wenjun Lu, Congli Li, Yongchang Shi, and Xiaoning Sun

New Star Research Institute of Applied Technology, Hefei 230031, China
{Wenjun.lu2013, shi756603491, sunxiaoning0117}@gmail.com,
lcliqa@163.com

Abstract. Image quality assessment (IQA) aims to provide computational models to measure the image quality consistently with subjective assessments. The SSIM index brings IQA from pixel-based to structure-based stage. In this paper, a new similarity index based on SIFT features (SIFT-SSIM) for full reference IQA is presented. In the algorithm, proportion of matched features in extracted features of reference image and structural similarity are combined into a comprehensive quality index. Experiments on LIVE database demonstrate that SIFT-SSIM is competitive with most of state-of-the-art FR-IQA metrics, and it can achieve higher consistency with the subjective assessments in some distortion types.

Keywords: Image Quality Assessment, Full Reference, Structural Similarity, Space Invariant Feature Transform.

1 Introduction

1.1 Image Quality Assessment

Image quality is an important indicator of varieties of image processing algorithms and optimizing system parameters. To establish an effective mechanism for image quality assessment has very important significance in image acquisition, encoding, network transmission and other areas [1]. In recent years, with the development of image processing technology, image quality assessment has attracted wide attention from researchers. There are many domestic and foreign research institutions and commercial companies join to research [2].

Image quality assessment can be divided into objective and subjective assessment methods, the former is by virtue of subjective perception experiments to evaluate the quality of an object; latter bases on model to give quantitative indicators, and to simulate perception mechanisms of human visual system to measure image quality. Relative to subjective quality assessment, objective quality assessment has become a research focus for advantages of simple, low cost, easy to parse and embed. Combination subjective with objective assessment method was focused on applications, and it uses results of subjective assessment to correct model parameters of objective quality assessment.

1.2 Full-Reference Methods Based on Structure Similarity

Because of applying HVS model to image quality assessment exist some problems, researchers have proposed image quality assessment model based on structural similarity. They consider that natural images have a specific structure, the pixels have a strong affiliation, and the affiliation reflects the structural information of visual scene. Zhou Wang and A.C. Bovik proposed image quality assessment method based on structural distortion, referred to SSIM [3]. In the method the comprehensive quality assessment model was proposed by using mean, standard deviation and unit standard deviation to represent brightness, contrast and structural similarity, it is the important landmark of IQA. The method deems that illumination is independent of object structure, and the light is mainly from changes in brightness and contrast. So it separates brightness and contrast from image structure information, and combines with structural information to image quality assessment. The method actually steers clear complexity of natural image content and multi-channel relations to evaluate structural similarity. It advantage to its lower complexity and wider applications. However, the algorithm only considers the structure of an image except to image features, but features of each image patches have different effects on image quality.

Subsequently, Zhou Wang and others also proposed a multi-scale SSIM [4], the algorithm obtained better results than a single scale. And the method introduced weights of information content to SSIM was proposed. In the algorithm, weights were calculated the proportion of information content of patches to the whole image in both reference and distorted image.

Lin Zhang et al proposed FSIM [5], two features of phase coherence and gradient were applied to calculate local similarity mapping. In pooling strategy of quality assessment, phase coherence was used again as a weighting function, because it can reflect local image in the perception of the importance of HVS well.

Lin Zhang et al also proposed RFSIM [6], in this method first-order and second-order Riesz transform were used to characterize the local structure of image, while Canny operator of edge detection was used to produces pooling mask of quality score.

Therefore, we propose combination strategy of feature matching and structural similarity, experiments approved that it can improve SSIM algorithm and even be comparable to other state-of-art full reference method.

2 Design of Full-Reference Algorithm Based on SIFT-SSIM

2.1 SIFT and SSIM

SIFT (Scale-invariant feature transform) algorithm was proposed by D.G.Lowe in 1999 [7], and in 2004 it was improved [8]. Later Y.Ke improved its partial descriptors by using PCA instead of histogram. SIFT algorithm is a local feature extraction algorithm, and it wants to find extreme points in scale space, extract location, scale and rotation invariant. SIFT features are local features of image, and remain invariant of rotation, scale, brightness, also maintain a certain stability of angle, affine transformation, and noise.

SIFT is based scale selection of image features, and it establishes multi-scale space, and detects same feature point at different scales, then determines the location of the feature point in their scales simultaneously determined to achieve the scale of anti-scaling purposes, bound by a number of points as well as low-contrast edge response points and rotation invariant feature descriptor extraction in order to achieve the purpose of anti-affine transformation. The algorithm mainly includes four steps:

- (1) To establish scale space, and to look for candidate points;
- (2) To confirm key points accurately, and to eliminate unstable points;
- (3) To acquire direction of key points;
- (4) To extract feature descriptors.

Results of SIFT feature extraction are shown in Figure 1.

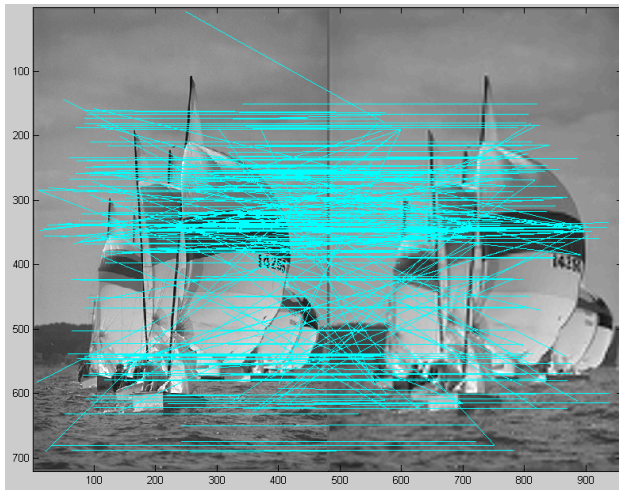


Fig. 1. Match Points of SIFT Features

Since SIFT has a good unique and rich amount of information for the reference image and degraded image feature matching between, it is particularly suitable for image quality assessment characteristic parameters. At the same time it has a scalable, can be very convenient feature vectors with other forms of joint, so consider using SIFT and structural similarity with the method to compensate for image features SSIM algorithm does not consider defects.

2.2 Framework of Improved Algorithm

The paper fusions SIFT features into comparison of luminance, contrast and structure, and a new metric of FR-IQA is presented. Diagram of the SIF-SSIM measurement system is shown in Figure 2.

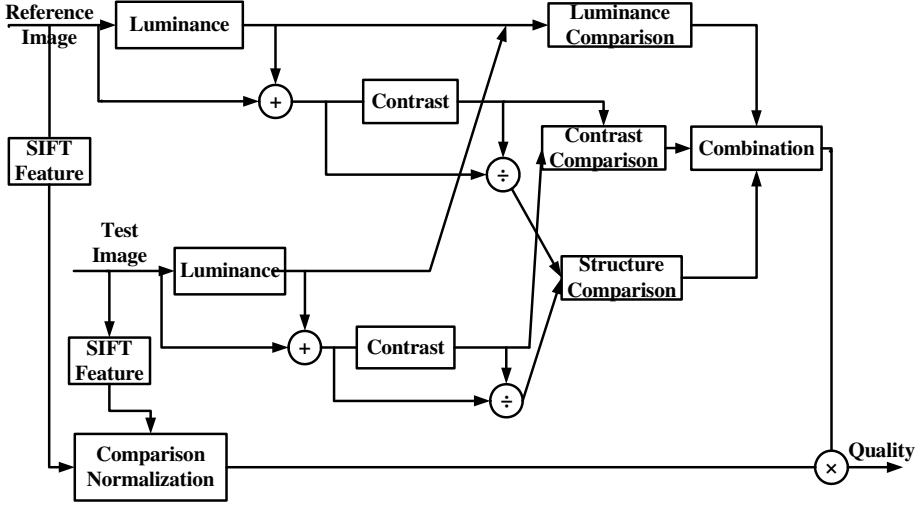


Fig. 2. Diagram of SIFT-SSIM Algorithm

When the two images SIFT feature vector generation, the next step we use the key feature vector images Euclidean distance as the two key points in determining the similarity measure. Take a key reference point in the image, and identify it with the test image Euclidean distance nearest first two key points in these two key points, if the closest distance divided by the distance is less than the proportion in threshold value, the acceptance of this pair of matching points. Lower this threshold ratio, SIFT matching points will reduce the number, but more stable.

In normalization process of comparison, we refer to proportion of SIFT matching points as one of quality assessment factors. Algorithm performs the following steps:

Algorithm: SIFT-SSIM.

- 1•To read a reference image x and a test image y .
- 2•To extract feature points of the two images, numbers of features are $NumSIFT(x)$ and $NumSIFT(y)$.
- 3•To Look for matching feature points•their numbers is $NumMatch(x, y)$.
- 4•To calculate the percentage to gain features score $SIFTScore = NumMatch(x, y) / NumSIFT(x)$.
- 5•To calculate brightness, contrast, and structure similarity functions of the two images, which are $l(x, y)$, $c(x, y)$, and $s(x, y)$, and to get SSIM score $SSIM(x, y) = l(x, y) \cdot c(x, y) \cdot s(x, y)$.
- 6•To calculate the final quality score $SIFT_SSIM = SIFTScore \cdot SSIM(x, y)$

3 Experiments Analysis and Comparison

To verify the performance of the algorithm, were carried out tests various distortion and overall tests on a public database of LIVE database2 [9]. Basic information of LIVE database is shown in Table 1.

Table 1. Basic Information of LIVE Image Database

Information/ Database	Number of Refer- ence Images	Numbers of Dis- torted Images	Number of Distortion	Image Type
LIVE	29	779	5	RGB
Distortion Type		Gaussian Blur JPEG JP2K White Noise Fast Fading		

Contrast algorithms include five classic full reference algorithms: PSNR [2], SSIM [3], MS-SSIM [4], FSIM [5], and RFSIM [6]. Index of comparison Algorithm is Spearman rank correlation coefficient (SROCC) and Pearson correlation coefficient (PLCC).

To find suitable threshold of quality assessment correlation, we range the Ratio from 0.1 to 0.9, and even form 0.91 to 0.99, then SROCC is calculated, which is shown in Table 2. The results show that, for the assessment of the role of the distortion effects in the two images, the high value of Ratio maybe helpful to enhance matching feature points. Therefore, the paper takes Ratio to 0.99.

Table 2. Relationship of Ratio and SROCC in FF

Ratio	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.91
	0.795	0.859	0.904	0.927	0.940	0.947	0.951	0.950	0.947
	4	3	2	4	6	6	7	4	6
Ratio	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	
	0.945	0.945	0.945	0.944	0.945	0.948	0.950	0.952	
	8	7	2	7	1	5	9	2	

The assessment results of various algorithms are shown in Tables 3 and 4.

Table 3. SROCC

Algorithm\Distortion	JP2K	JPEG	WN	Blur	FF	ALL
PSNR	0.9119	0.8774	0.9365	0.7669	0.8869	0.8759
SSIM	0.9370	0.8974	0.9099	0.9065	0.9396	0.9181
MS-SSIM	0.9370	0.8983	0.9213	0.9344	0.9350	0.9252
FSIM	0.9386	0.8965	0.9189	0.9527	0.9515	0.9316
RFSIM	0.9275	0.8890	0.9291	0.8914	0.9232	0.9120
SIFT-SSIM	0.9378	0.9060	0.9179	0.9298	0.9522	0.9287

Table 4. PLCC

Algorithm\Distortion	JP2K	JPEG	WN	Blur	FF	ALL
PSNR	0.9227	0.9020	0.9291	0.7736	0.8913	0.8837
SSIM	0.9295	0.9108	0.9059	0.9113	0.9446	0.9204
MS-SSIM	0.9230	0.9018	0.8363	0.7999	0.7896	0.8501
FSIM	0.9217	0.9034	0.8730	0.9604	0.9498	0.9217
RFSIM	0.9338	0.9091	0.9125	0.9037	0.9260	0.9170
SIFT-SSIM	0.9362	0.9180	0.9237	0.9155	0.9529	0.9293

In Tables above, we make the most prominent data bold. It is clear that the proposed algorithm in two distortions of JPEG and FF has achieved the best results, its overall performance exceeds to most algorithms, and it can be compared with FSIM.

4 Conclusion

Aims to ignorance features of the structural similarity algorithm, and while SIFT features have a variety of invariant and scalability, we consider to build comprehensive image quality assessment index with structural similarity and SIFT features. SIFT-SSIM of full reference algorithm was proposed, and experiments verify its high performance.

Acknowledgement. This work is supported by the Anhui Natural Science Foundation of China (Grant No. 1208085MF97).

References

1. Pang, L.-L., Li, C.-L., Luo, J.: Summary of Image Quality assessment Technology. *Avionics Technology* 42(2), 31–35 (2011)
2. VQEG. Final report from VQEG on the validation of objective models of video quality assessment (March 15, 2000), http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseII/downloads/VQEGII_Final_Report.pdf
3. Wang, Z., Alan, C.B., Hamid, R.S.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13(4), 600–612 (2004)
4. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multi-scale structural similarity for image quality assessment. In: *IEEE Asilomar Conference on Signals, Systems and Computers*, pp. 1398–1402 (2003)
5. Zhang, L., Zhang, L., Mou, X., Zhang, D.: FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing* 20(8), 2378–2386 (2011)
6. Zhang, L., Zhang, L., Mou, X.: RFSIM: a feature based image quality assessment metric using Riesz transforms. In: *Proc. IEEE International Conference on Image Processing*, pp. 321–324 (2010)

7. Lowe, D.G.: Object recognition from local scale-invariant features. In: International Conference on Computer Vision, pp. 1150–1157 (1999)
8. Lowe, D.G.: Distinctive image features from scale-invariant key points. *International Journal of Computer Vision* 60(2), 91–110 (2004)
9. Sheikh, H., Wang, Z., Cormack, L., Bovik, A.: Live image quality assessment database release 2 (2005)

Edge Detection in Presence of Impulse Noise

Yuying Shi¹, Feng Guo², Xinhua Su³, and Jing Xu⁴

¹ Department of Mathematics and Physics, North China Electric Power University, Beijing, 102206 China

² Department of Mathematics and Physics, North China Electric Power University, Beijing, 102206 China

³ School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan, China

⁴ School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou, China

Abstract. Edge detection in image processing is a difficult but meaningful problem. In this paper, we propose a variational model with L^1 -norm as the fidelity term based on the well-known Mumford-Shah functional. To solve it, we devise fast numerical algorithms through applying the binary label-set method. Numerical experiments on gray-scale images are given. By comparing with the famous Ambrosio-Tortorelli model with L^1 -norm as the fidelity term, we demonstrate that our model and algorithms show advantages in efficiency and accuracy for impulse noise.

Keywords: Mumford-Shah model, binary level set method, edge detection, split Bregman method.

1 Introduction

Edge detection shows how important it is in image processing, computer vision, and in many other fields, such as material science and physics [24,3]. According to the contexts of image processing, edge detection means extracting the boundaries of some objects from a given image. A lot of methods have been proposed for this purpose, such as gradient operator (Roberts operator, Sobel operator, Prewitt operator) (see, e.g., [16]), second-order derivatives (LOG operator, Canny operator) (see, e.g., [1]) and some new methods (using wavelets, fuzzy algorithms etc.) (see, e.g., [7]). The capability of using gradient operator is limited, as the accuracy of edge identification is usually deteriorated in presence of noise. Though the second-order derivative operators have advantages in denoising and smoothing the edge, they can blur the images which do not have noise. We also notice the recent researches have favor in using a variety of filter banks to improve the accuracy of edge detection, and the interested readers are referred to [4,13,12,18,22] and the references therein.

The seminal Mumford-Shah model [15] aims to simultaneously solve the problems of denoising and edge detection. The Mumford-Shah (MS) model is:

$$\min_{u, \Gamma} \left\{ E(u, \Gamma) = \mu \int_{\Omega \setminus \Gamma} |\nabla u|^2 d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} (u - I)^2 d\mathbf{x} + |\Gamma| \right\}, \quad (1)$$

where the minimizer u is intended to get the piecewise smooth approximation of a given image I on an open bounded domain $\Omega \subset \mathbb{R}^2$, Γ is the edge set, and μ, ν are positive tuning parameters. Many important properties have been obtained (see, e.g., [9,10]).

According to comparison with region-based image segmentation, edge detection also focuses on locating open curves belonging to constituent element of edges, yet do not have interior and exterior regional separation. In a recent conference report [25], we proposed to embed an open (or a closed) curve into a narrow region (or band), formed by the curve and its parallel curve (also known as the offset curve [23]). We then combined the Mumford-Shah (MS) model [15] with the binary leveling processing by introducing the binary level-set function: $\psi = 1$, if x is located in small regions around the edges, while $\psi = 0$ otherwise [25]. So the modified Mumford-Shah (MMS) model is to replace the length term in (1) with:

$$TV(\psi) = \sup_{\mathbf{p} \in S} \int_{\Omega} \psi \operatorname{div} \mathbf{p} \, d\mathbf{x}, \quad S := \left\{ \mathbf{p} \in C_c^1(\Omega; \mathbb{R}^2) : |\mathbf{p}| \leq 1 \right\}, \quad (2)$$

where $C_c^1(\Omega; \mathbb{R}^2)$ is the space of vector-valued functions compactly supported in Ω with first-order partial derivatives being continuous. However, Wang et al. [25] introduced a very preliminary algorithm for this modified model. Since the L^1 -norm has showed advantage in handling impulse noise (see, e.g., [14,11]), we borrowed the above form (2) and consider the modified MS model with L^1 -norm as the fidelity term:

$$\min_{\psi \in \{0,1\}, u} \left\{ \mu \int_{\Omega} (1 - \psi)^2 |\nabla u|^2 \, d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} |u - I| \, d\mathbf{x} + TV(\psi) \right\}. \quad (3)$$

It is clear that $\psi \in \{0,1\}$ in (3) is non-convex. So, we relax the set and constrain $\psi \in [0,1]$ (see, e.g., [5]) as follows:

$$\min_{\psi \in [0,1], u} \left\{ \mu \int_{\Omega} (1 - \psi)^2 |\nabla u|^2 \, d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} |u - I| \, d\mathbf{x} + TV(\psi) \right\}. \quad (4)$$

An auxiliary g can be used to handle the last term and face the constraint $g = u - I$ by a penalty method. Thus (4) can be approximated by the following problem with L^1 -norm:

$$\min_{\substack{u, g \\ \psi \in [0,1]}} \left\{ \mu \int_{\Omega} (1 - \psi)^2 |\nabla u|^2 \, d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} |g| \, d\mathbf{x} + \frac{\xi}{2} \int_{\Omega} |u - I - g|^2 \, d\mathbf{x} + TV(\psi) \right\}. \quad (5)$$

The rest of the paper is organized as follows. In Section 2, for solving the minimization problem, we separate it into several subproblems, and apply fixed-point iterative method and split Bregman method [10] to solve the u -subproblem and ψ -subproblem. In Section 3, we make a comparison with the seminar Ambrosio-Tortorelli model [2] with L^1 fidelity term and present numerical results to demonstrate the strengths of the proposed method.

2 The Minimization Algorithms

In this section, we introduce the minimization algorithms for solving (5) and use an iterative method for computing u , and the split Bregman method for resolving the binary level-set function ψ .

Similarly as [6,26], we use the alternating optimization technique to split (5) into three subproblems:

– u -subproblem: for fixed ψ and g , we solve

$$\min_u \left\{ \mu \int_{\Omega} (1 - \psi)^2 |\nabla u|^2 d\mathbf{x} + \frac{\xi}{2} \int_{\Omega} |u - I - g|^2 d\mathbf{x} \right\}. \quad (6)$$

– ψ -subproblem: for fixed u , we solve

$$\min_{\psi \in [0,1]} \left\{ \mu \int_{\Omega} (1 - \psi)^2 |\nabla u|^2 d\mathbf{x} + TV(\psi) \right\}. \quad (7)$$

– g -subproblem: for fixed u , we solve

$$\min_g \left\{ \frac{\nu}{2} \int_{\Omega} |g| d\mathbf{x} + \frac{\xi}{2} \int_{\Omega} (u - I - g)^2 d\mathbf{x} \right\}. \quad (8)$$

The solution of (8) can be easily expressed as:

$$g = |u - I| \max \left\{ 0, 1 - \frac{\nu}{2\xi|u - I|} \right\}.$$

The detailed process can be referred to [21,20]. Next, we present the algorithms for (6) and (7).

2.1 Fixed-Point Iterative Method for Solving u

We first consider (6). Notice that the functional in (6) is convex for fixed ψ , so it admits a minimizer. The corresponding Euler-Lagrange equation takes the form:

$$\begin{cases} -2\mu \operatorname{div}((1 - \psi)^2 \nabla u) + \xi(u - I - g) = 0, & \text{in } \Omega, \\ \frac{\partial u}{\partial \mathbf{n}} \Big|_{\partial\Omega} = 0, & \text{on } \partial\Omega. \end{cases} \quad (9)$$

where \mathbf{n} is the unit outer normal vector to $\partial\Omega$ as before. We expect ψ take value 0 at the homogeneous region, i.e., $1 - \psi \approx 1$. So, we propose to use a fixed-point iterative scheme based on relaxation method to solve this elliptic problem with variable coefficient (see, e.g., [17] for similar ideas). We start with the difference equation:

$$\begin{aligned} \xi u_{i,j} = & 2\mu[(1 - \psi)_{i,j+1}^2 (u_{i,j+1} - u_{i,j}) + (1 - \psi)_{i,j-1}^2 (u_{i,j-1} - u_{i,j}) \\ & + (1 - \psi)_{i-1,j}^2 (u_{i-1,j} - u_{i,j}) + (1 - \psi)_{i+1,j}^2 (u_{i+1,j} - u_{i,j})] + \xi I_{i,j} + \xi g_{i,j}, \end{aligned} \quad (10)$$

where $u_{i,j} \equiv u(i, j)$ is the approximate solution of (9) at grid point (i, j) with grid size $h = 1$ as usual. Then applying the Gauss-Seidel iteration to (10) leads to

$$\begin{aligned} & (2\mu(C_E + C_W + C_N + C_S) + \nu)u_{i,j}^{k+1} = \\ & 2\mu[C_E u_{i,j+1}^k + C_W u_{i,j-1}^{k+1} + C_N u_{i-1,j}^{k+1} + C_S u_{i+1,j}^k] + \xi I_{i,j} + \xi g_{i,j}, \end{aligned} \quad (11)$$

where

$$C_E = (1 - \psi^k)_{i,j+1}^2, \quad C_W = (1 - \psi^k)_{i,j-1}^2, \quad C_N = (1 - \psi^k)_{i-1,j}^2, \quad C_S = (1 - \psi^k)_{i+1,j}^2.$$

And we implement the relaxation method to speed up the iteration (11) by

$$u_{i,j}^{k+1} = u_{i,j}^k - \omega_1 r_{i,j}^{k+1}, \quad (12)$$

where $\omega_1 > 0$ is the relaxation factor. Collecting (12), we have the new concrete scheme:

$$u_{i,j}^{k+1} = \frac{u_{i,j}^k + \omega_1 (\xi I_{i,j} + \xi g_{i,j} + 2\mu[C_E u_{i,j+1}^k + C_W u_{i,j-1}^{k+1} + C_N u_{i-1,j}^{k+1} + C_S u_{i+1,j}^k])}{1 + \omega_1 (2\mu(C_E + C_W + C_N + C_S) + \xi)}. \quad (13)$$

2.2 Split Bregman method for solving ψ

Now, we will apply the split Bregman method [11] to solve (7). First, we define

$$\rho(\psi) := \mu \int_{\Omega} (1 - \psi)^2 |\nabla u|^2 d\mathbf{x}, \quad (14)$$

and

$$Tr(\psi) := \begin{cases} 1, & \psi > 1, \\ \psi, & 0 \leq \psi \leq 1, \\ 0, & \psi < 0. \end{cases} \quad (15)$$

to make sure that $\psi \in [0, 1]$ for every iteration [8]. Here, we introduce an auxiliary variable \mathbf{d} and solve the following problem:

$$\min_{\psi} \left\{ \int_{\Omega} |\mathbf{d}| d\mathbf{x} + \rho(\psi) \right\} \quad \text{subject to} \quad \mathbf{d} = (d_1, d_2) = \nabla \psi,$$

where $|\mathbf{d}| = \sqrt{d_1^2 + d_2^2}$ and $\rho(\psi)$ is defined in (14). Following the technique in [11], the split Bregman iteration can be formulated as

$$(\psi^{n+1}, \mathbf{d}^{n+1}) = \arg \min_{\psi, \mathbf{d}} \left\{ \int_{\Omega} |\mathbf{d}| d\mathbf{x} + \rho(\psi) + \frac{\lambda}{2} \int_{\Omega} (\mathbf{d} - \nabla \psi - \mathbf{b}^n)^2 d\mathbf{x} \right\}, \quad (16)$$

for given \mathbf{b}^n , and

$$\mathbf{b}^{n+1} = \mathbf{b}^n + (\nabla \psi^{n+1} - \mathbf{d}^{n+1}). \quad (17)$$

It is equivalent to write (16)-(17) with the form of $\mathbf{d} = (d_1, d_2)$ and $\mathbf{b} = (b_1, b_2)$.

$$\begin{aligned} (\psi^{n+1}, d_1^{n+1}, d_2^{n+1}) = \arg \min_{\psi, d_1, d_2} & \left\{ \int_{\Omega} \sqrt{d_1^2 + d_2^2} \, d\mathbf{x} \right. \\ & + \mu \int_{\Omega} (1 - \psi)^2 |\nabla u|^2 \, d\mathbf{x} + \frac{\lambda}{2} \int_{\Omega} (d_1 - \partial_x \psi - b_1^n)^2 \, d\mathbf{x} \\ & \left. + \frac{\lambda}{2} \int_{\Omega} (d_2 - \partial_y \psi - b_2^n)^2 \, d\mathbf{x} \right\}, \end{aligned} \quad (18)$$

and

$$b_1^{n+1} = b_1^n + \left(\partial_x \psi^{n+1} - d_1^{n+1} \right), \quad b_2^{n+1} = b_2^n + \left(\partial_y \psi^{n+1} - d_2^{n+1} \right).$$

The Euler-Lagrange equation of (18) for ψ with fixed d_1 and d_2 is

$$-\lambda \Delta \psi + 2\mu |\nabla u|^2 (\psi - 1) - \lambda \partial_x (d_1 - b_1^n) - \lambda \partial_y (d_2 - b_2^n) = 0,$$

with the Neumann boundary conditions $\frac{\partial \psi}{\partial \mathbf{n}} = 0$. Applying the same technique in Subsection 2.1, we need to solve the equations about ψ :

$$\psi_{i,j}^{k+1} = \frac{\psi^k + \omega_2 (F_{i,j}^k + \lambda (\psi_{i+1,j}^k + \psi_{i-1,j}^{k+1} + \psi_{i,j-1}^{k+1} + \psi_{i,j+1}^k))}{1 + \omega_2 (2\mu |\nabla u|_{i,j}^{k+1}|^2 + 4\lambda)}, \quad (19)$$

where $\omega_2 > 0$ is the relaxation factor and

$$F := 2\mu |\nabla u|^2 + \lambda \partial_x (d_1 - b_1^n) + \lambda \partial_y (d_2 - b_2^n).$$

For fixed ψ , the optimality condition of (18) with respect to d_1 and d_2 gives

$$d_1 = \max \left\{ h - \frac{1}{\lambda}, 0 \right\} \frac{h_1}{h}, \quad d_2 = \max \left\{ h - \frac{1}{\lambda}, 0 \right\} \frac{h_2}{h}, \quad (20)$$

where

$$h_1 = \partial_x \psi + b_1^n, \quad h_2 = \partial_y \psi + b_2^n, \quad h = \sqrt{h_1^2 + h_2^2}.$$

Now, we summarize the split Bregman (SB) algorithm as follows.

Split Bregman Algorithm

1. Initialization: set $d_1^0 = d_2^0 = b_1^0 = b_2^0 = 0$, and input $\psi^0, u^0, \mu, \nu, \omega_1, \omega_2, \lambda$.
2. For $n = 0, 1, \dots$,
 - (i) Update u^{n+1} using the iteration scheme (13) with initial value for iteration: u^n and ψ^n (in place of ψ);
 - (ii) Update g^n by

$$g^n = |u - I|^n \max \left\{ 0, 1 - \frac{1}{\xi |(u - I)^n|} \right\};$$

- (iii) Update ψ^{n+1} using the iteration scheme (19) with initial values ψ^n, u^{n+1} (in place of u), and enforce $\psi^{n+1} \in [0, 1]$ by (15);
- (iv) Update d_1^{n+1} and d_2^{n+1} by (20) with ψ^{n+1} in place of ψ ;
- (v) Update b_1^{n+1} and b_2^{n+1} by

$$b_1^{n+1} = b_1^n + (\partial_x \psi^{n+1} - d_1^{n+1}), \quad b_2^{n+1} = b_2^n + (\partial_y \psi^{n+1} - d_2^{n+1}).$$

3. Endfor till some stopping rule meets.

It is necessary to point out the iteration in Step 2 (i) can be ran for several times.

2.3 Algorithm for Ambrosio-Tortorelli Model

In this section, we present the new AT model equipped with L^1 -norm as the fidelity term, and provide numerical results to compare the relevant two algorithms: SB algorithm and AT algorithm which are shown as follows. The Ambrosio-Tortorelli model with L^1 -norm as the fidelity term is:

$$E_{AT}(u, v) = \mu \int_{\Omega} (v^2 + o_{\varepsilon}) |\nabla u|^2 d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} |u - I| d\mathbf{x} + \int_{\Omega} \left(\varepsilon |\nabla v|^2 + \frac{1}{4\varepsilon} (v - 1)^2 \right) d\mathbf{x}, \quad (21)$$

where ε is a sufficient small parameter, and o_{ε} is any non-negative infinitesimal quantity approaching 0 faster than ε . Shah [19] also considered replacing the first term of the above model (21) by L^1 -functions. We can refer [9] to find something else about the AT model with L^1 -norm.

Similarly, we set $g = u - I$ by a penalty method. Then equation (21) can be approximated as follows:

$$E_{AT}(u, v, g) = \mu \int_{\Omega} (v^2 + o_{\varepsilon}) |\nabla u|^2 d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} |g| d\mathbf{x} + \frac{\xi}{2} \int_{\Omega} |u - I - g|^2 d\mathbf{x} + \int_{\Omega} \left(\varepsilon |\nabla v|^2 + \frac{1}{4\varepsilon} (v - 1)^2 \right) d\mathbf{x}, \quad (22)$$

As mentioned above, we need solve the following subproblems of the AT model with L^1 fidelity term (22) :

$$\begin{cases} u = \frac{2\mu}{\xi} \operatorname{div}[(v^2 + o_{\varepsilon}) \nabla u] + I + g, & \text{in } \Omega, \\ v \left(\frac{1}{4\varepsilon} + \mu |\nabla u|^2 \right) = \varepsilon \Delta v + \frac{1}{4\varepsilon}, & \text{in } \Omega, \\ g = |u - I| \max \left\{ 0, 1 - \frac{\nu}{2\xi |u - I|} \right\}, & \text{in } \Omega, \\ \frac{\partial u}{\partial \mathbf{n}} = \frac{\partial v}{\partial \mathbf{n}} = 0, & \text{on } \partial\Omega. \end{cases} \quad (23)$$

Using the same relaxation technique as above, we solve the first u equation (23) by the iterative scheme:

$$u_{i,j}^{n+1} = \frac{u_{i,j}^n + \omega_3 (\xi I_{i,j} + \xi g_{i,j} + 2\mu [C_E u_{i,j+1}^n + C_W u_{i,j-1}^{n+1} + C_N u_{i-1,j}^{n+1} + C_S u_{i+1,j}^n])}{1 + \omega_3 (2\mu (C_E + C_W + C_N + C_S) + \xi)}, \quad (24)$$

where $\omega_3 > 0$ is the relaxation factor and

$$C_E = (v^n)_{i,j+1}^2 + o_\epsilon, C_W = (v^n)_{i,j-1}^2 + o_\epsilon, C_N = (v^n)_{i-1,j}^2 + o_\epsilon, C_S = (v^n)_{i+1,j}^2 + o_\epsilon.$$

Also, we solve v by the iterative scheme:

$$v_{i,j}^{n+1} = \frac{v_{i,j}^n + \omega_4 (\frac{1}{4\epsilon} + \epsilon (v_{i+1,j}^n + v_{i-1,j}^{n+1} + v_{i,j-1}^{n+1} + v_{i,j+1}^n))}{1 + \omega_4 (\frac{1}{4\epsilon} + 2\mu |\nabla u_{i,j}^{n+1}|^2 + 4\epsilon)}, \quad (25)$$

where $\omega_4 > 0$ is the relaxation factor.

Now, we present the full algorithm as follows.

AT Algorithm

1. Initialization: set $o_\epsilon = \epsilon^{\hat{p}}$ ($\hat{p} > 1$), and input $u^0, v^0, g^0, \mu, \nu, \xi, \hat{p}, \epsilon, \omega_3, \omega_4$.
2. For $n = 0, 1, \dots$,
 - (i) Update u^{n+1} by (24);
 - (ii) Update v^{n+1} by (25);
 - (iii) Update g by the third equation in (23);
3. Endfor till some stopping rule meets.

3 Numerical Experiments

In this section, we compare the above two algorithms: the SB algorithm and the AT algorithm. Here, we set the stopping rule by using the relative error

$$E(u^{n+1}, u^n) := \|u^{n+1} - u^n\|^2 \leq \eta, \quad (26)$$

where a prescribed tolerance $\eta > 0$. The choice of the parameters are specified in the captions of the figures.

In the first experiment, we test three real clean images (see Fig. 1), and generally choose the same parameters in the SB algorithm and the AT algorithm with

$$\mu = 5 \times 10^3, \quad \nu = 1, \quad \omega_1 = 1 \times 10^{-3}, \quad \eta = 1 \times 10^{-6}$$

and the parameters $\xi = 300, \omega_2 = 1, \lambda = 500$ in the SB algorithm and $\epsilon = 2 \times 10^{-3}, \hat{p} = 2$ in the AT algorithm, respectively. For simplicity, we let $\omega_3 = \omega_4 = \omega_2$ in AT algorithm. These parameters are chosen by experimental experiences to get good edges in visual. We present the input images and the results detected

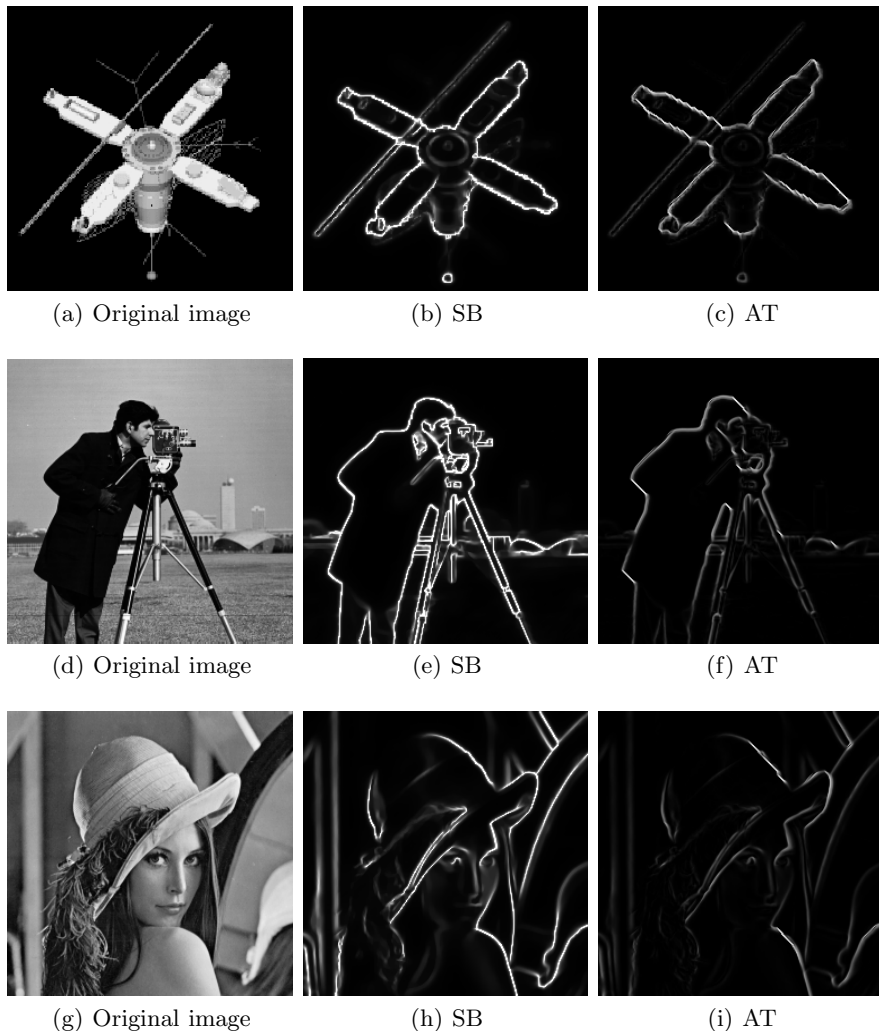


Fig. 1. Comparisons of clean images. Column 1: original images; Column 2: detected edges by the SB algorithm; Column 3: detected edges by the AT algorithm.

by two different algorithms in Fig. 1. We observe from the above figures that the SB algorithm outperforms the AT algorithm. And the SB algorithm is able to detect all the meaningful edges.

In the second experiments, we turn to the comparison of two algorithms for the noisy images in Fig. 2 (salt-pepper noise with $\sigma = 0.04$). In this situation, the u equation usually needs more than one inner iteration in order to smooth the noisy image. We choose $\lambda = 1 \times 10^3$ and other parameters are the same as

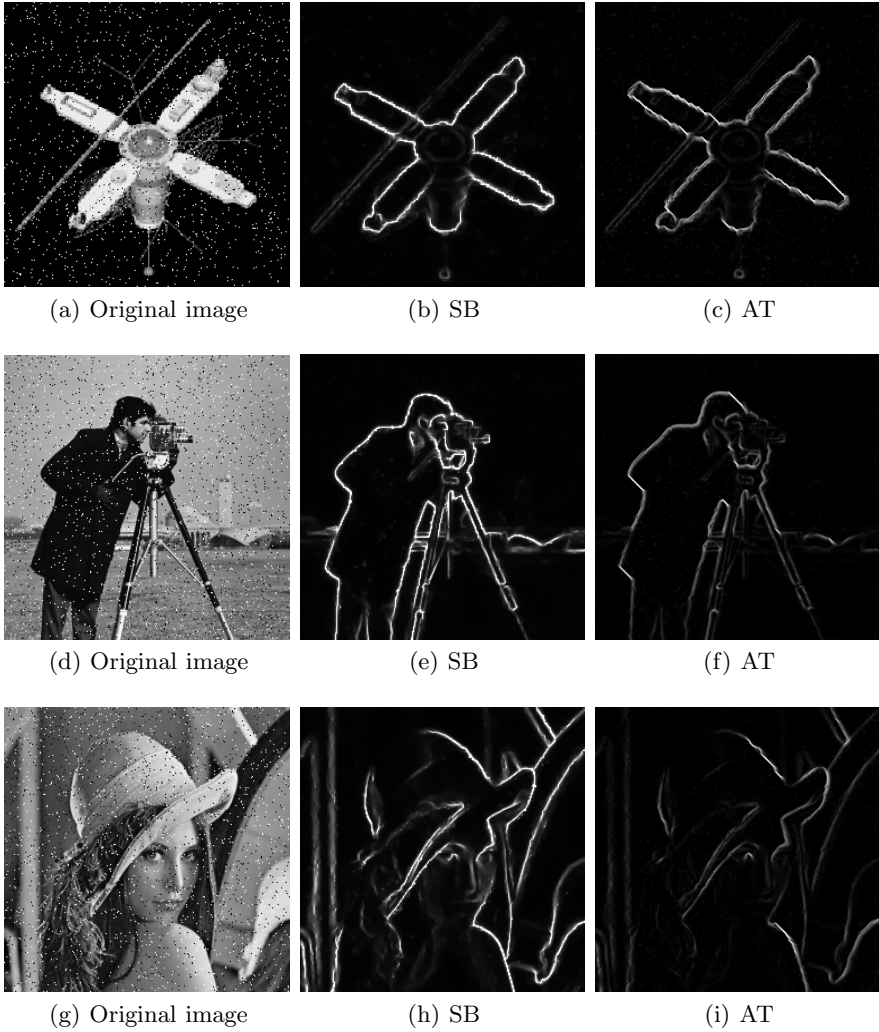


Fig. 2. Comparisons of noisy images corrupted by salt and pepper noise with $\sigma = 0.04$. Column 1: original images with salt-pepper noise; Column 2: detected edges by the SB algorithm; Column 3: detected edges by the AT algorithm.

the previous example. The original images and the detected results obtained by the SB algorithm and the AT algorithm are shown in Fig. 2. Obviously, the SB algorithm yields better results and the AT algorithm smooths some details of the edges.

4 Concluding Remarks

We first present a modified Mumford-shah model with L^1 fidelity term by introducing one L^1 -norm term to approximate the edge length, and show an efficient algorithm based on the split Bregman iteration to solve the minimum problem. Comparing with the seminar Ambrosio-Tortorelli model which introduces a quadratic integral of an edge signature to approximate the edge length, we design the gradient descent algorithm. Numerical experimental results show that our approximation of the edge length are robust, and our algorithm based on split Bregman iteration are effective and accurate.

Acknowledgments. This research supported by NSFC (No. 11271126) and the Fundamental Research Funds for the Central Universities.

References

1. Alvarez, L., Lions, P., Morel, J.: Image selective smoothing and edge detection by nonlinear diffusion. ii. *SIAM J. Numer. Anal.* 29(3), 845–866 (1992)
2. Ambrosio, L., Tortorelli, V.: Approximation of functions depending on jumps by elliptic functions via Γ -convergence. *Comm. Pure Appl. Math.* 13, 999–1036 (1990)
3. Berkels, B., Rätz, A., Rumpf, M., Voigt, A.: Extracting grain boundaries and macroscopic deformations from images on atomic scale. *J. Sci. Comput.* 35(1), 1–23 (2008)
4. Brook, A., Kimmel, R., Sochen, N.: Variational restoration and edge detection for color images. *J. Math. Imaging Vis.* 18(3), 247–268 (2003)
5. Brown, E., Chan, T., Bresson, X.: A convex relaxation method for a class of vector-valued minimization problems with applications to Mumford-Shah segmentation. *UCLA cam report cam 10–44*, pp. 10–43 (2010)
6. Cai, J., Osher, S., Shen, Z.: Split Bregman methods and frame based image restoration. *Multiscale Model. Sim.* 8(2), 337–369 (2009)
7. Catté, F., Lions, P., Morel, J., Coll, T.: Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal.* 29(1), 182–193 (1992)
8. Chan, R., Tao, M., Yuan, X.: Constrained total variation deblurring models and fast algorithms based on alternating direction method of multipliers. *SIAM J. Imaging Sci.* 6(1), 680–697 (2013)
9. Erdem, E., Sancar-Yilmaz, A., Tari, S.: Mumford-shah regularizer with spatial coherence. In: Sgallari, F., Murli, A., Paragios, N. (eds.) *SSVM 2007*. LNCS, vol. 4485, pp. 545–555. Springer, Heidelberg (2007)
10. Erdem, E., Tari, S.: Mumford-shah regularizer with contextual feedback. *J. Math. Imaging. Vis.* 33(1), 67–84 (2009)
11. Goldstein, T., Osher, S.: The split Bregman method for L1 regularized problems. *SIAM J. Imaging Sci.* 2(2), 323–343 (2009)
12. Llanas, B., Lantarón, S.: Edge detection by adaptive splitting. *J. Sci. Comput.* 46(3), 486–518 (2011)
13. Meinhardt, E., Zacur, E., Frangi, A., Caselles, V.: 3D edge detection by selection of level surface patches. *J. Math. Imaging Vis.* 34(1), 1–16 (2009)

14. Micchelli, C., Shen, L., Xu, Y., Zeng, X.: Proximity algorithms for image models ii: L1/tv denoising. *Adv. Comput. Math.* (2011)
15. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.* 42(5), 577–685 (1989)
16. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE T. Pattern Anal.* 12(7), 629–639 (1990)
17. Perona, P., Malik, J.: Scale-space and edge-detection using anisotropic diffusion. *IEEE T. Pattern Anal.* 12(7), 629–639 (1990)
18. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the Mumford-Shah functional. In: 12th International Conference in Computer Vision, pp. 1133–1140. IEEE (2009)
19. Shah, J.: A common framework for curve evolution, segmentation and anisotropic diffusion. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 136–142 (1996)
20. Shi, Y., Wang, L., Tai, X.: Geometry of total variation regularized Lp-model. *J. Comput. Appl. Math.* 236(8), 2223–2234 (2012)
21. Tai, X., Wu, C.: Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model. In: Tai, X.-C., Mørken, K., Lysaker, M., Lie, K.-A. (eds.) *SSVM 2009*. LNCS, vol. 5567, pp. 502–513. Springer, Heidelberg (2009)
22. Tao, W., Chang, F., Liu, L., Jin, H., Wang, T.: Interactively multiphase image segmentation based on variational formulation and graph cuts. *Pattern Recogn.* 43(10), 3208–3218 (2010)
23. Toponogov, V.: *Differential geometry of curves and surfaces: A concise guide*. Birkhauser (2006)
24. Upmanyu, M., Smith, R., Srolovitz, D.: Atomistic simulation of curvature driven grain boundary migration. *Interface Sci.* 6, 41–58 (1998)
25. Wang, L.-L., Shi, Y., Tai, X.-C.: Robust edge detection using Mumford-Shah model and binary level set method. In: Bruckstein, A.M., ter Haar Romeny, B.M., Bronstein, A.M., Bronstein, M.M. (eds.) *SSVM 2011*. LNCS, vol. 6667, pp. 291–301. Springer, Heidelberg (2012)
26. Wu, C., Tai, X.: Augmented lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models. *SIAM J. Imaging Sci.* 3, 300–339 (2010)

A Novel Multi-focus Image Capture and Fusion System for Macro Photography

Yili Zhao^{1,3}, Yi Zhou^{2,3}, and Dan Xu³

¹ Southwest Forestry University, Kunming Yunnan 650224

² Yunnan Normal University, Kunming Yunnan 650091

³ Yunnan University, Kunming Yunnan 650091

ylzhao@swfu.edu.cn, zhouyikm@gmail.com, danxu@ynu.edu.cn

Abstract. This paper proposes a novel multi-focus image capture and fusion system for macro photography. The system consists of three components. The first component is a novel multi-focus image capture device which can capture multiple macro images taken at different focus distances from a photographic subject, with high precision. The second component is a feature based method which can align multiple in-focus images automatically. The third component is a new multi-focus image fusion method which can combine multiple macro images to a fused image with a greater depth of field. The proposed image fusion method is based on Gaussian and Laplacian pyramids with a novel weight map selection strategy. Several data sets are captured and fused by the proposed system to verify the hardware and software design. Subjective and objective methods are also used to evaluate the proposed system. By analyzing the experimental results, it shows that this system is flexible and efficient, and the quality of the fused image is comparable to the results of other methods.

Keywords: Macro photography, multi-focus image, image alignment, Laplacian pyramid, image fusion.

1 Introduction

Macro photography is gaining popularity and has been used in more and more fields like zoology, botany and film production [1]. It is suitable for human visual perception and computer-processing tasks such as segmentation, feature extraction and object recognition. Traditionally, professional macro lens are used in macro photography. However, due to the limited depth-of-field of optical lenses in the CCD devices, it is often not possible to get an image that contains all relevant objects in focus. Consequently, the obtained image will not be in focus everywhere, i.e., if one object in the scene is in focus, another one will be out of focus. In order to obtain all objects in focus, multiple photos with different focus region are necessary and multi-focus fusion is used to blend these images to a fused image.

In general, users can set the initial distance between the lens and the object, and adjust the lens focal length manually. However, this is tedious and error-prone. The paper proposes a novel macro photography capture device which can capture a serial

of macro images taken at different focus distances from a photographic subject, with high precision. After multiple photos are captured, user also needs to fuse these photos to a sharp one. Multi-scale transforms are often used to analyze the information content of images for the purpose of fusion, and various methods based on multi-scale transforms have been proposed in literatures, such as Laplacian pyramid based methods, gradient pyramid based methods and discrete wavelet based methods [2-8].

The basic idea of multi-scale transform is to perform a multi-resolution decomposition on each source image, then integrate all these decomposition coefficients to produce a composite representation. The fused image is finally reconstructed by performing an inverse transform. The conventional wavelet-based methods consider the maximal absolute value of wavelet coefficients or local feature of two images [9-11]. Wavelets are very effective in representing objects with isolated point singularities, while wavelet bases are not the most significant in representing objects with singularities along lines. As a consequence, the method based on the wavelet cannot excavate the edge quality and detail information [12]. The paper proposes a multi-focus image fusion method based on Laplacian pyramid with a novel weight map selection strategy.

Most multi-focus fusion algorithms for macro photography are assumed that all source images are point-wise correspondence, that is, the colors at and around any given pixel in one image correspond to the colors at and around that same pixel in another image. However, when using the mechanical device to capture different in-focus images, small motion between adjacent images is inevitable. In these cases, these images need to be aligned very well before fusion. Otherwise, motion blur effect will appear in the final fused image. This paper proposes a feature based image alignment strategy to register the captured images before multi-focus image fusion.

The rest of this paper is organized as follows. The multi-focus image capture device design is discussed in section 2. The multi-focus image alignment method is described in section 3. The multi-focus image fusion based on Laplacian pyramids with weighted maps is given in section 4. After that, experimental results analysis and evaluation are proposed in section 5. This section also illustrates the proposed macro photography capture and fusion system with some practical data samples. Finally, the last section gives some concluding remarks.

2 Macro Photography Capture Device Design

An auto-controlled image capture device is developed to obtain the macro photos for further processing as shown in figure 1. The device consists of a DSLR Camera with macro lens; a slide platform to hold the camera moving back and forth; a screw rod to push the platform; a stepper motor as mechanical power, and a MCU-based control unit. The camera shutter, flash, and stepper motor are attached to the control unit to implement the auto-photography system. The thread pitch of the screw rod is 4mm and with a 360 degree spin, thus the rod pushes the platform to move 4mm. The step angle of the stepper is 1.8 degree and 8 subdivision controlling scheme is used. Therefore, with one control pulse, the stepper motor spins for 0.225 degree and the platform moves 0.0025mm. In the experiments, the focal length of the camera is fixed.

The control unit generates 40 pulses to push the camera for 0.1mm, and a signal is sent to the shutter and flash to capture an image. This procedure is repeated for many times until the displacement of the first and last image has covered the field length of requirement. Consequently, a series of images with same subject but different focus point is obtained.

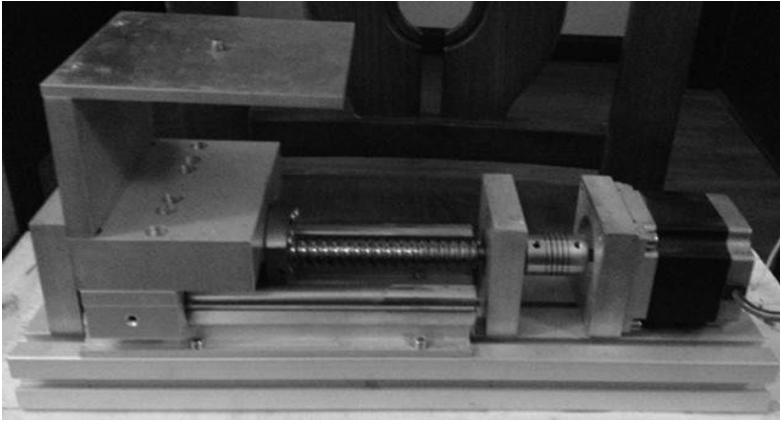


Fig. 1. Macro photo capture device

3 Multi-focus Image Alignment

Currently, most multi-focus fusion algorithms are assumed that source images are point-wise correspondence, that is, the colors at and around any given pixel in one image correspond to the colors at and around that same pixel in another image. However, when using the mechanical device to capture different in-focus images, small motion between adjacent images is inevitable. In these cases, the input images must be aligned very well before fusion.

In order to register multiple in-focus images and calculate their motion models, this paper first extracts MOPS feature from every input image. The MOPS algorithm proposed by Matthew Brown [13] is a relatively lightweight scale invariant feature detector compared with SIFT algorithm [14], and has the advantage of faster detection speed. The MOPS algorithm extended Harris algorithm with rotation and scale invariance. When matching the feature points between different in-focus images, it is necessary to perform nearest neighbor search in the feature space. Using kd-tree based nearest neighbor search algorithm can reduce search time complexity. The fast feature matching algorithm based on k-d tree is described as:

- (a) Segmenting foreground and background for each source image;
- (b) Detecting MOPS for each source image's foreground;
- (c) Constructing kd-tree for each image's feature point set;
- (d) Traversing every feature points of each image, initial image index $i = 0$, and feature point index $n = 0$. For the n^{th} feature point of image i , find the

nearest neighbor nn_1 and second nearest neighbor nn_2 to all other images, and their Euclidean distance d_1 and d_2 . If the value of d_1 / d_2 is less than 0.6, nn_1 is considered as the best match point;

- (e) Once all image feature points have been traversed, it is also need to validate the feature matching results. Assume the feature point n_j of image I has matching index n_{jy} with image J , then checks whether the matching index of the feature point n_{jy} of image J with image I equals to n_j ;
- (f) Using RANSAC algorithm [15] to estimate the motion model between adjacent images, and exclude the outliers.

4 Multi-focus Image Fusion

Through the above registration process, a series of images with good alignment can be obtained, and these aligned images are the input of the multi-focus fusion algorithm. Multi-focus fusion needs to compute the desired image by keeping the “best” parts in the multi-focus image sequence. This process is guided by the quality measure, which will be consolidated into a scalar-valued weight map. It is useful to think of the multi-focus image sequence as a stack of images. The final image is then obtained by collapsing the image stack using weighted blending strategy. In this step, it is assumed that images are perfectly aligned. Otherwise, the input sequence should be aligned first as in part 3.

One of the most effective and canonical method used to describe image with multi-resolution is the image pyramid proposed by Burt and Adelson [16]. The essential idea of image pyramid is to decompose the source image into different spatial resolutions through some mathematical operations. The most commonly used two image pyramids representation are Gaussian pyramid and Laplacian pyramid, and Laplacian pyramid can be derived from the Gaussian pyramid, which is a multi-scale representation obtained through a recursive low-pass filtering and down-sampling operations. For example, the finest level G_0 of Gaussian pyramid is equal to the original image, and the next level G_1 is obtained by recursively down-sampling G_0 and so on. Both sampling density and resolution are decreased from level to level of the pyramid, and this local averaging process which generates each pyramid level from its predecessor is called REDUCE:

$$G_{+1} = REDUCE(G) \quad (1)$$

$$G_{+1} = \sum_{m=-2}^2 \sum_{n=-2}^2 w(m, n) G(2i + m, 2j + n) \quad (2)$$

where $\mathcal{W}(m, n)$ is a two-dimensional separable Gaussian window function.

The Gaussian pyramid is a set of low-pass filtered images. In order to obtain the band-pass images required for the multi-resolution spline it is necessary to subtract each level of the pyramid from the next lowest level. Let \hat{G}_{+1} be the image obtained by expanding G_{+1} , and \hat{G}_{+1} has the same size as G . Then the EXPAND operator can be defined as:

$$\hat{G}_{+1} = \text{EXPAND}(G_{+1}) \quad (3)$$

$$\hat{G}_{+1} = 4 \sum_{m=-2}^2 \sum_{n=-2}^2 \mathcal{W}(m, n) G_{+1} \left(\frac{i+m}{2}, \frac{j+n}{2} \right) \quad (4)$$

Therefore, the Laplacian pyramid can be defined as:

$$\begin{cases} L_l = G - \hat{G}_{+1}, & 0 \leq l < N \\ L_N = G_N, & l = N \end{cases} \quad (5)$$

where N is the number of Laplacian pyramid levels.

An important property of the Laplacian pyramid is that it is a complete image representation. The steps used to construct the pyramid may be reversed to recover the original image exactly. Therefore, the reconstruction of original image from the Laplacian pyramid can be expressed as:

$$\begin{aligned} G_0 &= L_0 + \text{EXPAND}(G) = L_0 + \text{EXPAND}(L_1 + \text{EXPAND}(G_2)) \\ &= L_0 + \text{EXPAND}(L_1 + \text{EXPAND}(L_2 + \dots + \text{EXPAND}(L_N))) \end{aligned} \quad (6)$$

First each image is converted to grayscale. Then each grayscale image is processed by a Laplacian filter. The absolute value of the filter response is calculated. The method then computes a weighted average along each pixel, using weights computed from the quality measure. To obtain a consistent result, it is needed to normalize the values of the N weight maps such that they sum to one at each pixel (i, j) :

$$\hat{W}_{j,k} = \left[\sum_{k'=1}^N W_{j,k'} \right]^{-1} W_{j,k} \quad (7)$$

where subscript ij , k refers to pixel (i, j) in the k^{th} image.

The resulting image R can then be obtained by a weighted blending of the input images:

$$R_j = \sum_{k=1}^N \hat{W}_{j,k} I_{j,k} \tag{8}$$

where I_k the k^{th} input image in the sequence.

However, applying equation (8) directly will produce an unsatisfactory result. Wherever weights change quickly, disturbing seams will appear. To address the seam problem, this paper uses a technique inspired by Burt and Adelson [16]. Their original technique seamlessly blends input images guided by an alpha mask, and works at multi-resolutions using pyramidal image decomposition. This technique is adapted to our case, where there are N images and N normalized weight maps that act as alpha mask. Let the l^{th} level in a Laplacian pyramid decomposition of an image A be defined as $\mathcal{L}A^l$, and $\mathcal{G}B^l$ for a Gaussian pyramid of image B . Then, the coefficients are fused in a similar fashion to equation (8):

$$\mathcal{L}R_{ij}^l = \sum_{k=1}^N \mathcal{G}W_{ij,k}^l \mathcal{L}I_{ij,k}^l \tag{9}$$

For example, each level l of the Laplacian pyramid is computed as a weighted average of the original Laplacian decomposition for level l , with the l^{th} level of Gaussian pyramid of the weight map serving as the weights. Finally, the pyramid $\mathcal{L}R^l$ is collapsed to obtain R . An overview of this framework is given in figure 2. For dealing with color images, each color channel will be fused separately.

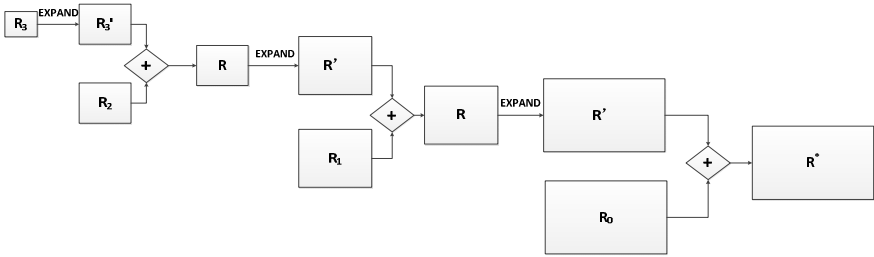


Fig. 2. Pyramid collapsing

5 Experiments Analysis and Evaluation

In order to test the proposed multi-focus capture and fusion system, several data sets have been captured by the device. Each data set has different focuses and its resolution is 2144x1424 pixels. The proposed fusion method is compared with other multi-focus fusion methods such as average method and wavelet method [17-19]. In the experiments, standard deviation, information entropy and average gradient are used to

evaluate the multi-focus image fusion result objectively. Corresponding experimental results are shown in following figures.

The first data set for evaluation consists of 61 images for a mantis. Three images of them are showed in figure 3(a), 3(b) and 3(c). Every image in the data set has different focus region as the small oval indicates. The fused image generated by the proposed method is shown in figure 3(d). It can be seen that those out-of-focus regions in three images are all in-focus in figure 3(d).

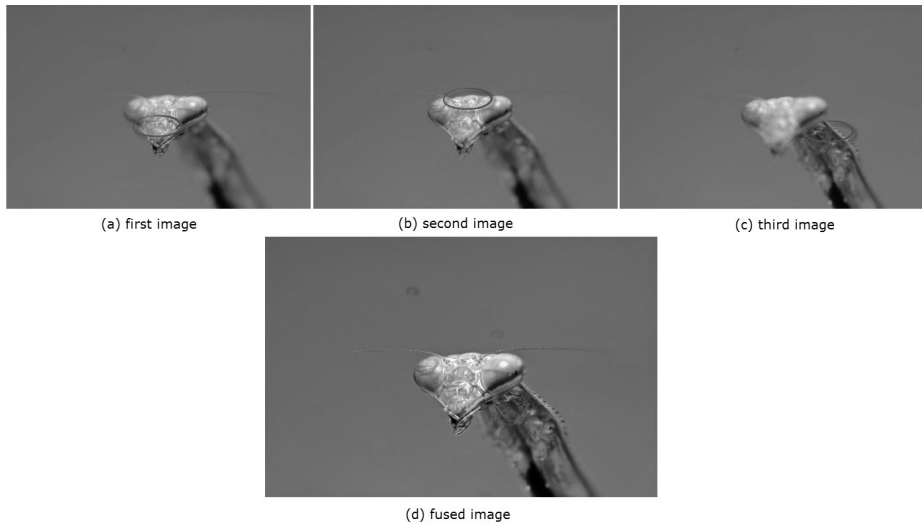


Fig. 3. Three images of the first data set and fused result

For comparison, the fusion results generated by average fusion, wavelet fusion and pyramid fusion are shown in figure 4.

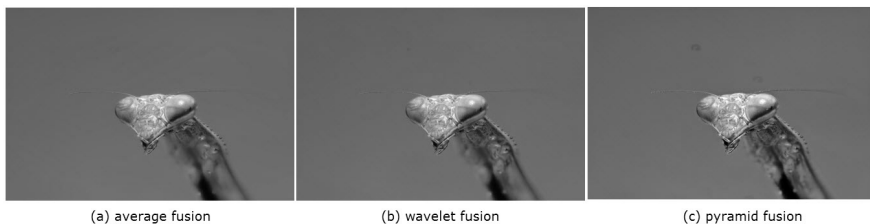


Fig. 4. Fusion results generated by different methods

Table 1 presents a quantitative comparison of various multi-focus fusion methods for this data set. The evaluation order is figure 4(a), 4(b) and 4(c). These fusion methods consists of average based method, wavelet based method and pyramid based method in terms of standard deviation, information entropy and average gradient. From table 1, wavelet based method has maximum value in terms of standard deviation, and pyramid based method has maximum values in terms of other two columns.

Table 1. Quantitative evaluation for the first data set

Standard deviation	Information entropy	Average gradient
55.4012	4.8904	0.0002767
55.6480	4.9104	0.0003714
55.3374	5.0465	0.0005607

The second data set for evaluation consists of 81 images for a bee. Six images of them are showed in figure 5. Every image in the data set has different focus region. The fusion results generated by three methods are shown in figure 6.

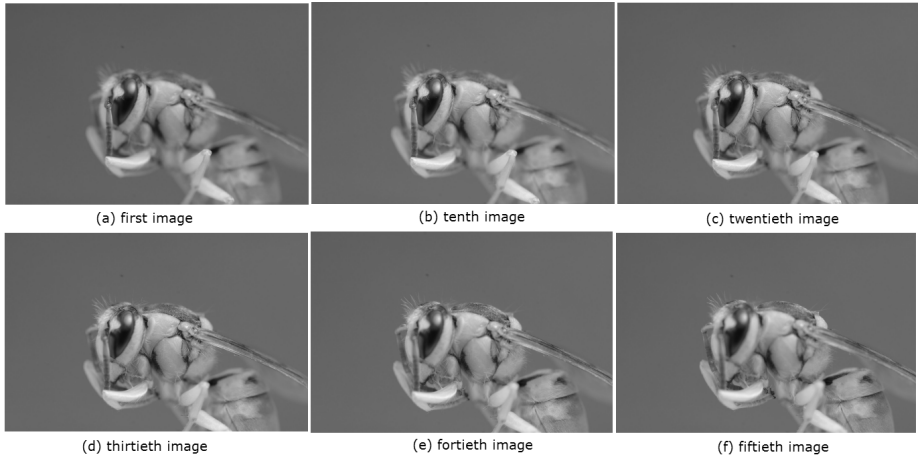
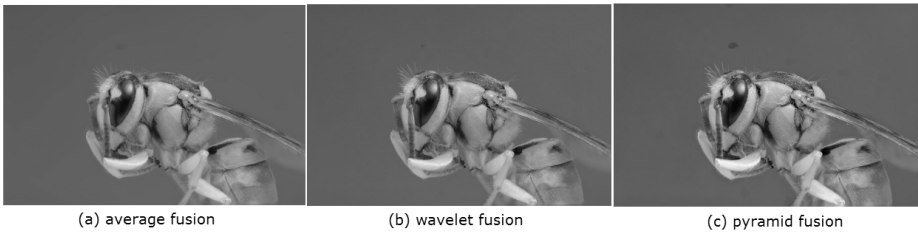
**Fig. 5.** Six images of the second data set**Fig. 6.** Fusion results generated by different methods

Table 2 presents a quantitative comparison of various multi-focus fusion methods for this data set. The evaluation order is figure 6(a), 6(b) and 6(c). These fusion methods consists of average based method, wavelet based method and pyramid based method in terms of standard deviation, information entropy and average gradient. From table 2, pyramid based method has maximum values in all three columns, which indicates that it can generate best fusion result.

Table 2. Quantitative evaluation for the second data set

Standard deviation	Information entropy	Average gradient
47.6883	5.3884	0.0006481
47.9520	5.4691	0.0008176
49.2818	5.6032	0.0010000

6 Conclusion

In this paper, a novel multi-focus image capture and fusion system for macro photography is proposed. The hardware component can capture multiple in-focus images with high precision. The software component can align a sequence of multi-focus images and fuse them. The proposed multi-focus fusion method is based on the Gaussian and Laplacian pyramids with weight maps extension. In order to make the evaluation of image quality more effective and more comprehensive, subjective and objective evaluations are adopted. Experiments demonstrate that the proposed system is flexible and efficient for macro photography capture and fusion.

References

1. Antonio, M.: Digital macro photography of cactus and succulent plants. *Cactus and Succulent Journal* 85, 101–106 (2013)
2. Burt, P., Kolczynski, R.: Enhanced image capture through fusion. In: *Proceedings of 4th International Conference on Computer Vision*, Berlin, pp. 173–182 (1993)
3. Wang, H., Jin, Z., Li, J.: Research and Development of Multiresolution Image Fusion. *Control Theory and Applications* 21, 145–149 (2004)
4. Jin, H., Yang, X., Jiao, L.: Image Enhancement via Fusion Based on Laplacian Pyramid Directional Filter Banks. In: *Proceedings of International Conference on Image Analysis and Recognition*, Toronto, pp. 239–246 (2005)
5. Haghghat, M., Aghagolzadeh, A., Seyedarabi, H.: Multi-Focus Image Fusion for Visual Sensor Networks in DCT Domain. *Computers and Electrical Engineering* 37, 789–797 (2011)
6. Haghghat, M., Aghagolzadeh, A., Seyedarabi, H.: Real-time fusion of multi-focus images for visual sensor networks. In: *Proceedings of 6th Iranian Machine Vision and Image Processing*, pp. 1–6. IEEE Press, New York (2010)
7. Pu, T., Ni, G.: Contrast-based image fusion using the discrete wavelet transform. *Optical Engineering* 39, 2075–2082 (2000)
8. Wen, Y., Li, Y.: The Image Fusion Method Based on Wavelet Transform in Auto-analysis of Pashm. *Journal of Sichuan University* 37, 36–40 (2000)
9. Wang, H., Peng, J., Wu, W.: Remote Sensing Image Fusion Using Wavelet Packet Transform. *Journal of Image and Graphics* 9, 922–937 (2002)
10. Wang, H., Jing, Z., Li, J.: Image fusion using non-separable wavelet frame. *Chinese Optics Letters* 9, 523–552 (2003)
11. Long, G., Xiao, L., Chen, X.: Overview of the applications of Curvelet transform in image processing. *Journal of Computer Research and Development* 2, 1331–1337 (2005)

12. Hassan, M., Shah, S.: Block Level Multi-Focus Image Fusion Using Wavelet Transform. In: Proceedings of the 2009 International Conference on Signal Acquisition and Processing, pp. 213–216. IEEE Press, Washington (2009)
13. Brown, M., Szeliski, R., Winder, S.: Multi-Image Matching using Multi-Scale Oriented Patches. In: Proceedings of International Conference on Computer Vision and Pattern Recognition, San Diego, pp. 510–517 (2005)
14. David, G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 91–110 (2004)
15. Fischler, M., Bolles, R.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communication of the ACM* 24, 381–395 (1981)
16. Burt, P., Adelson, E.H.: A Multiresolution Spline with Application to Image Mosaics. *ACM Transactions on Graphics* 2, 217–236 (1983)
17. Valdimir, S., Petrović, C., Xydeas, S.: Gradient-Based Multi-resolution Image Fusion. *IEEE Transactions on Image Processing* 3, 228–236 (2004)
18. Liu, Z., Tsukada, K., Hanasaki, K.: Image fusion by using steerable pyramid. *Pattern Recognition Letters* 22, 929–939 (2001)
19. Li, S., Yang, B.: Multifocus image fusion using region segmentation and spatial frequency. *Image and Vision Computing* 26, 971–979 (2008)

Comparative Study of Near-Lossless Compression by JPEG XR, JPEG 2000, and H.264 on 4K Video Sequences

Wang Dan

China Academy of Space Technology, Beijing 100094, China
wangdan05@gmail.com

Abstract. Lossless or near-lossless compression for high resolution image is required to meet the increasingly high quality demands in various application fields. With respect to this requirement, ultra-high definition image/video compression technology is researched in this paper. We first delve into the JPEG XR compression algorithm and parameters, then report a comparative study evaluating rate-distortion performance between JPEG XR, H.264 and JPEG 2000. A set of five sequences with resolution of 4K(3840x2160) have been used. The Result shows that, for the test sequences used to carry out the experiments, the JPEG XR outperforms other two coding standards in terms of the trade off between compression efficiency and hardware complexity.

Keywords: JPEG XR, JPEG 2000, H.264, 4K.

1 Introduction

In most application fields, the image/video quality is the higher the better from the user's perspective, however, due to the limited channel and storage resources, the image/video has to be compressed. To meet the high quality requirement, there is an urgent need to find the proper video compression standard and tools for high resolution video sequences, with considering the tradeoff between coding efficiency and hardware complexity, to achieve the lossless or near-lossless compression result.

For this need, we first introduce the new international compression standard JPEG XR in Section 2, where we also made some exploration and research of JPEG XR tools. In Section 3, a comparative study evaluating rate-distortion performance between JPEG XR, H.264 and JPEG 2000 and the experiment result is given. Finally we conclude our work in Section 4.

2 Study on JPEG XR

2.1 JPEG XR Introduction

JPEG XR [1] is an international image coding standard, based on HD Photo developed by Microsoft technology. It is designed for the high dynamic range (HDR) and the high definition (HD) photo size. The XR of JPEG XR means the extended range.

The goal of JPEG XR is to achieve state-of-the-art image compression, while keeping the encoder and the decoder complexity lower [2]. It supports high compression performance twice as high as JPEG, and also has an advantage over JPEG2000 in terms of computational cost [3]. Motion JPEG XR [4] Standard specifies the use of JPEG XR coding for timed sequences of images

The main difference between JPEG XR, JPEG and JPEG 2000 is shown in following table.

Table 1. Comparison between JPEG, JPEG 2000 and JPEG XR

Standard	JPEG	JPEG 2000	JPEG XR
HDR support	×	√	√
Subjective lossless compression ratio	≈1/5	≈1/10	≈1/10
Lossy/lossless compression	Lossy (JPEG LS support lossless)	Lossy/lossless	Lossy/lossless
Scalability	×	best	Good
Key technology	Discrete Cosine Transform (DCT), Huffman entropy coding	Discrete Wavelet Transform(DWT), adaptive entropy coding	Photo Core Transform(PCT), adaptive entropy coding
Encoder complexity	1(base)	15 times	8 times

Like JPEG 2000, an image is divided into small parts so that they can be processed one by one more easily in JPEG XR standard. As in Figure 1, three hierarchies are presented to divide an image. They are tiles, macroblocks and blocks.

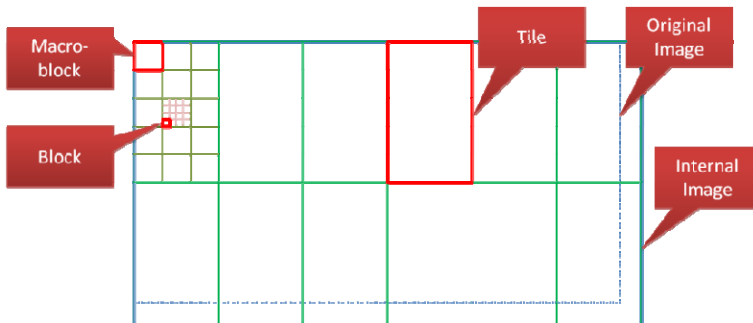


Fig. 1. An example of hierarchical image partitioning in JPEG XR

Tile size has to be decided at the beginning of JPEG XR compression since it is largely related to the hardware architecture [5]. The influence of tile size on coding efficiency is explored in Section 2.2

Macroblock is a basic data unit in JPEG XR. The size of a macroblock is predefined by the standard. One tile consists of $N \times M$ Macroblocks and one macroblock consists of 4×4 blocks which consists of 4×4 pixels.

The data flow of JPEG XR encoding is shown in Fig. 2.

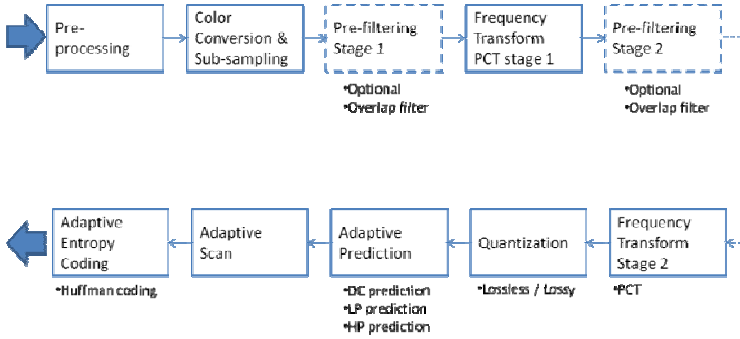


Fig. 2. JPEG XR Encoding Diagram

First of all, pre-processing is applied to the input image for better coding efficiency. Then the color conversion module that converts the original image color to another color space which is better suited for down-sampling is applied.

Then a transform called photo core transform (PCT) is applied to decompose an image into frequency components. The PCT is applied to a rectangular area called a macroblock. A transform called photo overlap transform (POT) is used with the PCT to eliminate the artifacts in the boundary between blocks. The influence of overlap filter on coding efficiency is studied in Section 2.2

Next, the transformed coefficients are quantized. Quantization parameters (QP) give a great impact on the image quality. After the coefficient prediction, an adaptive scanning is processed and coefficients are rearranged from two-dimensional form to one dimensional form.

Finally, the scanned coefficients are entropy coded using adaptive Huffman tables. Different from JPEG, JPEG XR encodes DC, LP, HP frequency bands separately. Not all values are entropy encoded. A process called normalization is used to decide which part of the bits should be kept as plain or be encoded. The plain bits in the DC and LP bands are called refinement while in the HP band they are called flexbits as shown in Fig. 3.

Flexbits is sent uncoded. For lossless 8 bit compression, Flexbits may account for more than 50% of the total bits. Flexbits forms an enhancement layer which may be omitted or truncated, where the parameter is called “TrimFlexBits”.

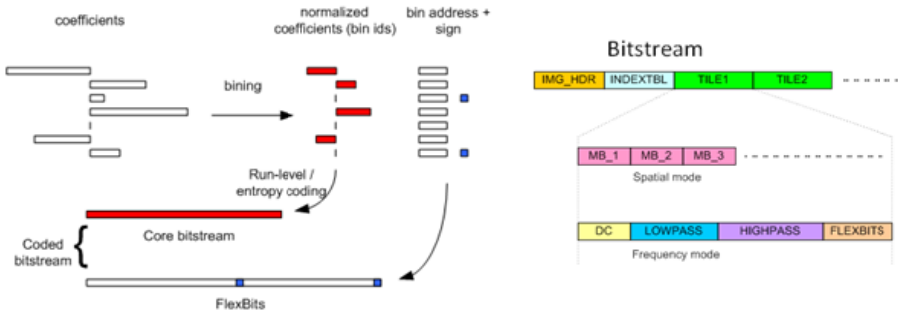


Fig. 3. FlexBits in JPEG XR

2.2 JPEG XR Evaluation

Experiments are performed to evaluate the influence of the main parameters on coding efficiency.

Input.

Five pictures (YUV444, 8bit per channel) extracted from five 4K video sequences is used for input in this evaluation shown in Fig. 4.

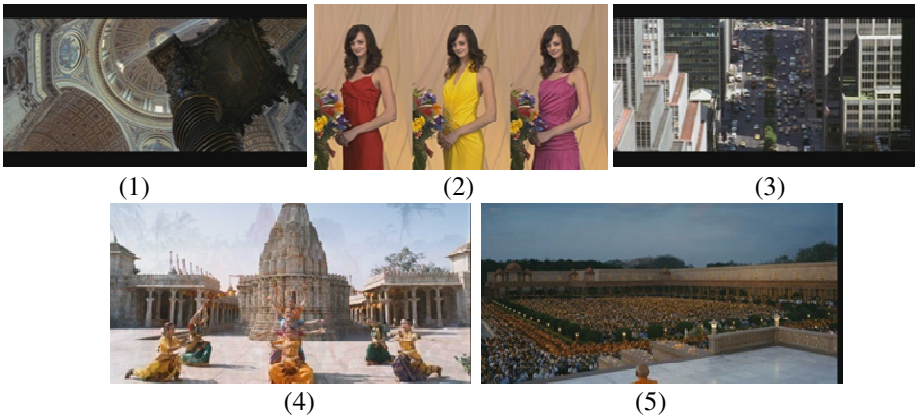


Fig. 4. Five 4K pictures for experiment

Experimental Results.

- Tile size

Test condition:

- Overlapfilter=1, Trimflexbits=0, these two parameters will be explained later in this chapter.
- TileSize=128x128, 512x512 and 3840x2160.

The rate-distortion result in terms of PSNR on the Y component is shown in the following graph.

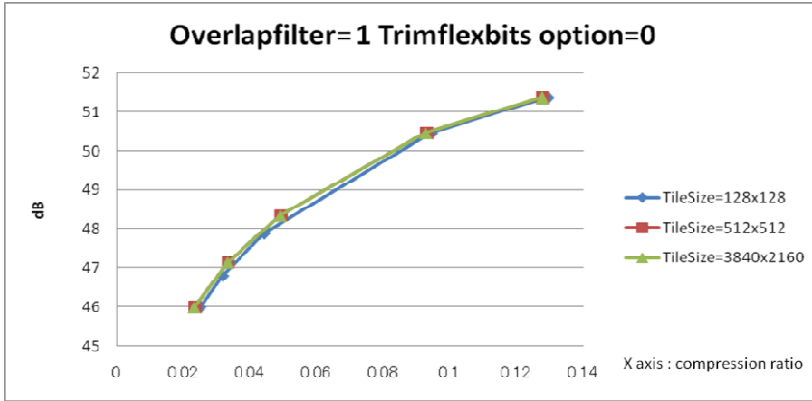


Fig. 5. Rate-distortion performance using PSNR for parameter TileSize

We can see from Fig. 5 that 512x512 and 3840x2160 perform better than 128x128 and the PSNR gain is among 0.1~0.3db, whereas there shows almost no difference between 512x512 and 3840x2160. The same result is also shown in Chroma components (U, V). Totally speaking, the influence of tile size on PSNR result is small but the influence of tile size on the memory size is large [5].

- Overlap filter (OF)

There are 3 values to be selected:

- 0: Overlap filter is not used neither in first step nor in second step
- 1: Overlap filter is used in first step
- 2: Overlap filter is used both in first step and in second step

The rate-distortion result in terms of PSNR and SSIM on both Y and UV component is shown in Fig. 6 and Fig. 7.

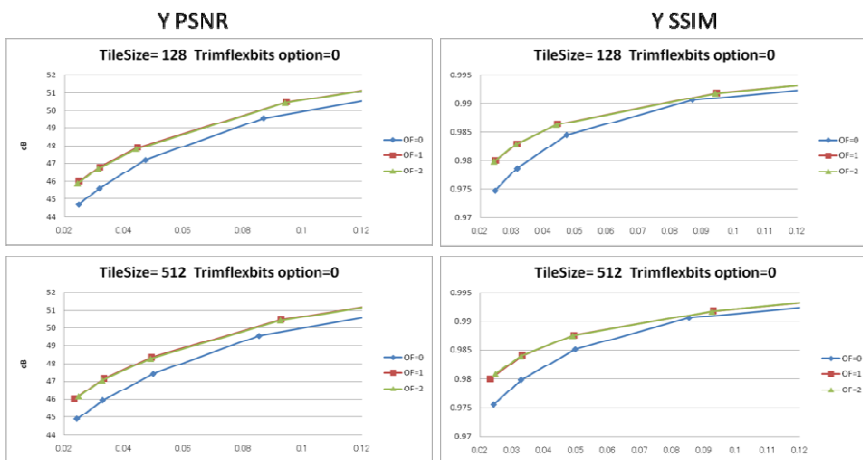


Fig. 6. Rate-distortion performance using PSNR and SSIM for parameter OF on Y component

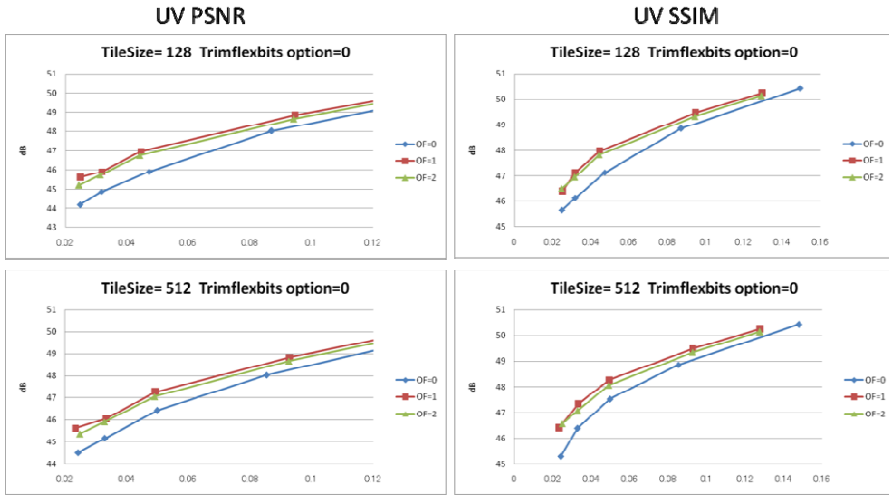


Fig. 7. Rate-distortion performance using PSNR and SSIM for parameter OF on UV component

As shown above, at the condition of TileSize=128x128, 512x512 and Trimflexbit= 0, following conclusions can be deduced:

- For both Y and UV components, OF=1 outperforms OF=0 with PSNR gain of 0.5~1.5dB
- For Y component, there is no obvious difference between OF=2 and OF=1
- For UV components, OF=2 outperform OF=1 with an average maximum gain of 0.2dB

The SSIM [6] result is similar with the PSNR result.

• TrimFlexBits

TrimFlexBits=[0,15], it means lower N bits will be truncated. If TrimFlexBits=0, then all FlexBits is preserved and if TrimFlexBits=15, all FlexBits is truncated.

The rate-distortion result in terms of PSNR on both Y and U component is shown in Fig. 8.

As shown above, at the condition of TileSize=512x512 and Overlapfilter= 0 and 1, the result shows that:

- At high compression ratio (below 0.05), the difference between the Trimflexbit option is very small
- At low compression ratio(over 0.1), the difference become larger, but the tendency seems different in Y and U(V)

• Overall performance

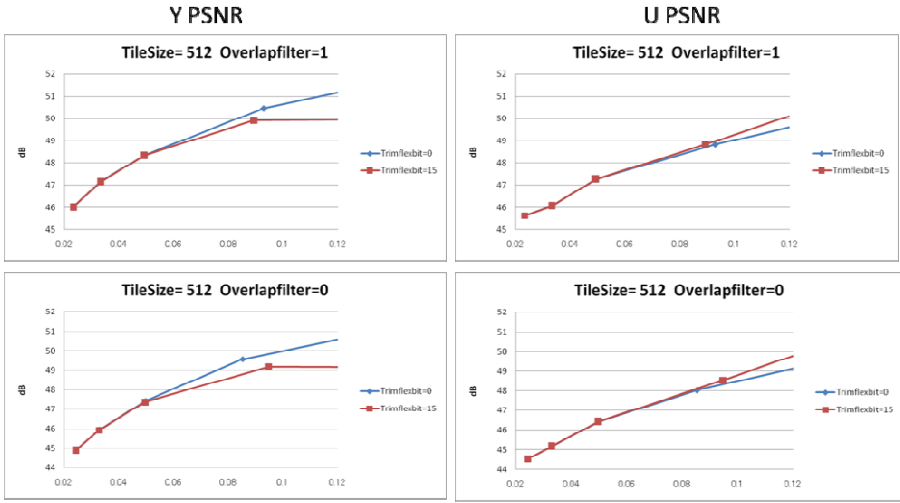


Fig. 8. Rate-distortion performance on Y and U component for Trimflexbit

Finally, we compared 5 profiles in terms of rate-distortion performance and complexity. The results are shown in Table 2 and Fig. 9.

Table 2. Definition and VTune cycles of 5 Profiles

	Overlap Filter	Tile Size	Trimflexbits	VTune ¹ [M Cycle]
Profile 1	0	128x128	15	5270.6
Profile 2	0	512x512	15	5212.2
Profile 3	1	512x512	0	6613.0
Profile 4	1	3840x2160	0	6589.2
Profile 5	2	512x512	0	6689.9

¹ Environment: Genuine Intel(R) CPU 2160 @ 1.80GHz 1.80GHz, 3.00 GB RAM, Intel(R) VTune(TM) Performance Analyzer 9.1

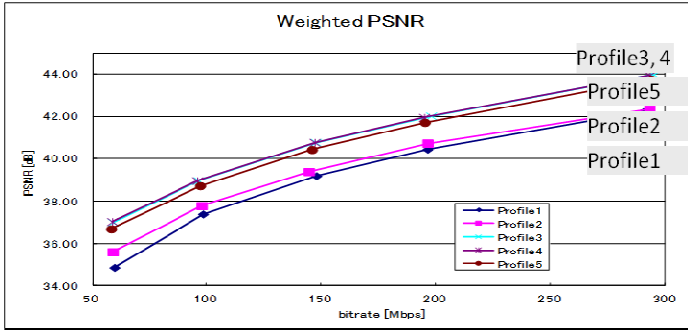


Fig. 9. Rate-distortion performance using weighted PSNR of 5 profiles

YUV Weight matrix is $\{0.7, 0.15, 0.15\}$ for weighted PSNR.

As shown above, profile 3 and profile 4 outperforms other profiles in terms of weighted PSNR at any bitrates, while the computation complexity of profile 3 and profile 4 are 25% larger than profile 1 and profile 2

3 Evaluation of 4K Video Compression Using JPEG XR, JPEG 2000 and H.264

For high quality compression for 4K video sequences, we evaluated 3 state-of-art encoding standard, including JPEG XR, JPEG 2000 and H.264.

Input

- 5 video sequences (3840x2160, YUV 444, 8 bits per channel, 30fps) have been shown in Fig. 4.
- Original bit-rate is $3840 \times 2160 \times 3 \times 8 \times 30 = 5972$ Mbps

Output

Bit rate: 60~300Mbps, Compression ratio: 1%~5%

Encoding Parameters

- JPEG XR

As evaluated in section 2, we chose appropriate parameters in terms of PSNR and complexity:

- Tile size: 512x512
- Overlap filter: 0 and 1
- Trimflexbits: 0

- JPEG 2000

Official software Jasper is used. Parameters are set as:

- Tile size: 3840x2160
- Code-block size: 64
- DWT Level: 4
- Rate: 0.03, 0.05, 0.07, 0.1, 0.15

- H.264

4 profiles are chosen, including AVC inter, IPP, IP and Intra. Parameters are set as the following table.

Table 3. Parameter of H.264/AVC encoder

	Inter	IPP	IP	Intra
ProfileIDC	100	100	100	100
IntraProfile	0	0	0	0
LevelIDC	50	50	50	50
IntraPeriod	15	3	2	1
IDRPeriod	0	0	0	1
NumberBFrames	1	0	0	0
RestrictSearchRange	2 (no restriction)	2	2	2
RDOptimization	0	0	0	0
RateControlEnable	0	0	0	0
RCUpdateMode	3	3	3	1 (original RC)

Experimental Results

The rate-distortion result in terms of average PSNR of 5 video sequences is shown in Fig 10. 7 encoding methods is compared in Table 4, including AVC inter, AVC intra, AVC IPP, AVC IP, JPEG 2000, JPEG XR.OF0 and JPEG XR OF1.

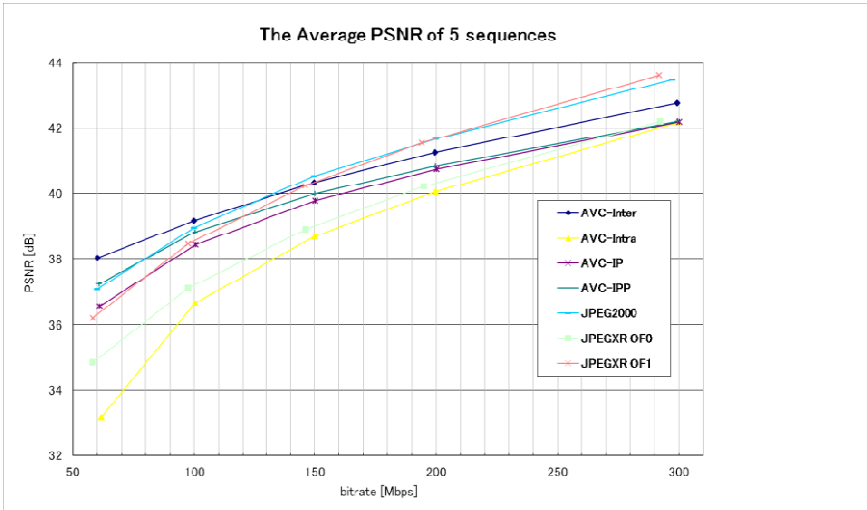


Fig. 10. Average PSNR of 5 sequences

Table 4. Bit-rate at 40dB of 7 coding methods

	Inter/Intra	Bit-rate [Mbps] at 40dB (Near-lossless coding)
AVC Inter(N=15)	Inter	136
AVC-IPP	Inter	150
AVC-IP	Inter	160
AVC-Intra	Intra	200
JPEG 2000	Intra	133
JPEG XR (OF=1)	Intra	140
JPEG XR (OF=0)	Intra	187

In the condition of high quality compression, it can be seen from Fig. 10 that:

- At 40dB point, the top 3 methods, including AVC Inter, JPEG 2000 and JPEG XR (OF=1), provide comparable compression efficiency for the conditions and the test material used to carry out the experiments, like the result in [7].
- Above the 40dB point, the AVC inter, JPEG 2000 and JPEG XR OF1 outperform others obviously
- JPEG XR OF1 outperforms AVC inter over 40.3 dB and outperforms JPEG 2000 over 41.5dB

However rate-distortion performance is not the only criteria for us to choosing the encoding standard, the hardware complexity is also very important. Take the complexity and resources requirement into account, the H.264 and JPEG 2000 is far more complex than JPEG XR [8,9]

4 Conclusion

In this paper, we firstly introduced JPEG XR standard, then we deeply evaluated the influence of three parameters on rate-distortion curve, including TileSize, OverlapFilter and Trimflexbits. It is found that TileSize=512x512, OverlapFilter=1 and Trimflexbits=0 is best combination for the high compression efficiency of 4K images.

After that, a simple evaluation methodology has been used to compare the compression performance between JPEG XR, JPEG 2000 and AVC inter and intra profile for compression of 4K video sequences. The result of our study shows that the AVC inter (N=15), JPEG 2000 and JPEG XR (OF=1) provide comparable compression efficiency in condition of high quality compression. The advantages of JPEG XR become more obvious with the higher PSNR. Meanwhile JPEG XR exhibits significantly lower computational demands and hardware complexity, so it is the most appropriate standard for the need of the near-lossless compression of 4K video.

References

1. JPEG XR Image Coding Specification, <http://www.itu.int/rec/T-REC-T.832>
2. JPEG XR, http://en.wikipedia.org/wiki/JPEG_XR
3. Jadhav, S.S., Jadhav, S.K.: JPEG XR an Image Coding Standard. *International Journal of Computer & Electrical Engineering* 4(2) (2012)
4. ITU-T Rec. T.833 (09/2010) JPEG XR Motion Format, <http://www.itu.int/rec/T-REC-T.833-201009-I/en>
5. Pan, C.H., Chien, C.Y., Chao, W.M., et al.: Architecture design of full HD JPEG XR encoder for digital photography applications. *IEEE Transactions on Consumer Electronics* 54(3), 963–971 (2008)
6. SSIM, http://en.wikipedia.org/wiki/Structural_similarity
7. De Simone, F., Ouaret, M., Dufaux, F., et al.: A comparative study of JPEG2000, H. 264, and HD photo. In: *Optical Engineering+ Applications*. International Society for Optics and Photonics, pp. 669602–669602-12 (2007)
8. Chien, C.Y., Huang, S.C., Pan, C.H., et al.: Full HD JPEG XR Encoder Design for Digital Photography Applications. Tech. (February 2010)
9. Marpe, D., George, V., Cycon, H.L., et al.: Performance evaluation of Motion-JPEG2000 in comparison with H. 264/AVC operated in pure intracoding mode. In: *Photonics Technologies for Robotics, Automation, and Manufacturing*. International Society for Optics and Photonics, pp. 129–137 (2004)

A Retinex-Based Local Tone Mapping Algorithm Using L_0 Smoothing Filter

Lei Tan, Xiaolin Liu^{*}, and Kaichuang Xue

College of Mechatronic Engineering and Automation
National University of Defense Technology
Changsha, Hunan, 410073, P.R. China
{tanlei08, lxlchangsha, xkcwork}@sina.com

Abstract. In this paper, we propose a novel halo-free local tone mapping algorithm using L_0 smoothing filter. Our method imitates the adaptation of the mechanism of the human visual system (HVS), which ensures a strong adaptability to the scenes of different dynamic ranges. Firstly, we will apply a global histogram adjustment method to the luminance image, which is a simple initial global adaptation; secondly, we will demonstrate how the luminance image is remapped by a retinex-based local tone mapping method. During the estimation of illumination, a L_0 smoothing filter is used instead of Gaussian filter to compress the contrast while reducing the halo artifacts; and finally, through the color correction, we will show how the tone-mapped RGB image is obtained. According to our experimental results, the proposed method outperforms the state-of-art tone mapping algorithms in color rendition and detail preservation.

Keywords: High dynamic range, tone mapping, retinex, L_0 smoothing filter, color correction.

1 Introduction

Dynamic range is defined by the luminance ratio of the highest scene luminance to the lowest, that is, the number of levels is divided in an image from the darkest grayscale to the brightest. An image is said to be at a High Dynamic Range (HDR) when its dynamic range exceeds by far of the display device. It's understood that an image is at the Low Dynamic Range (LDR) when its dynamic range is below the display device. Tone mapping algorithm is a method that compresses the dynamic range of the HDR images so that the mapped image can fit into the dynamic range of the display devices.

There are many tone mapping algorithms which can be divided into two categories: global and local tone mapping operators. The global tone mapping operators are computationally very simple since all the pixels of the original image are mapped using the same transformation. But one should be aware that the global tone mapping operators may cause the loss of details in the dark and bright areas of HDR image when the dynamic range of the image is too large. The local tone mapping operators

^{*} Corresponding author.

are thus necessary to better preserve the details and local contrast in images. However, local tone mapping operators are at a higher cost, and may cause halo artifacts because the local tone mapping operators consider the local information in the mapping processing for each individual pixel.

It doesn't matter whether the scene is dark or bright, the Human Visual System (HVS) can rapidly adapt to the luminance of the scene; therefore the HVS is the best tone mapping operator [1]. It is important to indicate that the HVS consists of a global and a local adaption. We can apply the HVS model to our tone mapping algorithm.

We have proposed a local tone mapping algorithm in this paper which reflects well the HVS. Our method helps to solve the problems encountered when applying the previously mentioned tone mapping algorithms, namely it does not produce halo artifacts and provides good color rendition. The specifics of the proposed method will be described in details below. In order to provide a better understanding, we will first demonstrate how a global tone mapping operator is applied to the luminance image, which imitates the initial adaption of the HVS; second, we will explain how to apply the retinex-based local tone mapping operator to compress the contrast while preserving the details; and finally, through the color correction, we will show the achievement of the tone-mapped image.

2 Previous Works

2.1 Retinex Algorithm

The retinex theory, first proposed by Land [2], intends to explain how the human visual system extracts reliable information from the world despite changes of illumination. Many image processing experiments [3] show that the retinex theory is consistent with lightness-color constancy theory. In other words, the color of objects is independent with illumination, and it only relies on the surface reflectance of objects.

In this paper, we will use retinex theory to solve the problem of high dynamic range compression. According to the retinex theory, the given image $I(x, y)$ satisfies formula (1):

$$I(x, y) = R(x, y) \times L(x, y) \quad (1)$$

in which $R(x, y)$ is the reflectance image, and $L(x, y)$ is the illumination image. The reflectance image R includes all the details of the original image, corresponding to the high-frequency components, while the illumination image L corresponds to the low-frequency components.

In order to compress the dynamic range of images, we should eliminate the effect of uneven illumination. Therefore, for calculating detail image R , illumination image L needs to be estimated. Then, subtracting illumination from the original image in the log-domain, gets dynamic-range-compressed image. The retinex algorithm [3] is given by

$$R(x, y) = \log I(x, y) - \log(F(x, y) * I(x, y)) \quad (2)$$

where $R(x, y)$ is the dynamic-range-compressed image; $I(x, y)$ is the original image, $*$ denotes the convolution operation; and $F(x, y)$ is the surround function.

In classical retinex algorithm, the surround function is Gaussian function. A drawback of the algorithm is that it induces halo artifacts along high-contrast edges. Under the assumption that the illuminant is spatially smooth, halo artifacts are due to the proximity of two areas of very different intensity.

2.2 L_0 Smoothing Filter

L_0 smoothing filter is an edge-preserving smoothing filter first proposed by Li Xu [4]. In order to sharpen the major edges of image while eliminate the low-amplitude structures, the filter calculates in an optimization framework using L_0 gradient minimization. The optimization framework controls the number of edges through globally controlling the number of non-zero gradients. The optimization framework can be expressed as the following formula:

$$\min_S \left\{ \sum_p (S_p - I_p)^2 + \lambda \cdot C(S) \right\} \quad (3)$$

$$C(S) = \text{count} \left\{ p \mid \left| \partial_x S_p \right| + \left| \partial_y S_p \right| \neq 0 \right\} \quad (4)$$

in which p is a pixel in the image; I is the original image; and S is the result image filtered by the L_0 smoothing filter. The gradient $\nabla S_p = (\partial_x S_p + \partial_y S_p)^T$ for each pixel p is calculated between neighboring pixels along the x and y directions. $C(S)$ counts p whose gradient-magnitude $\left| \partial_x S_p \right| + \left| \partial_y S_p \right|$ is not zero. λ is a smoothing parameter, and a large λ makes the result image having few edges.

Compared with current edge-preserving smoothing filters depending on local features like Bilateral filter (BLF) [5] and weighted least square (WLS) [6] filter, L_0 smoothing filter manages to locate important and prominent edges globally. Fig. 1 shows the corresponding performance comparison. Obviously, L_0 smoothing filter is the best edge-preserving smoothing filter of the three filters.

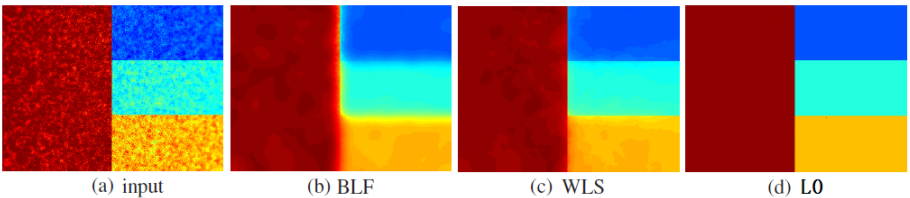


Fig. 1. Smoothing performance comparison. From (a) to (d): noisy input, results of BLF, WLS filter and L_0 smoothing filter

3 Proposed Method

The classic retinex algorithm has some drawbacks as described in section 2. These drawbacks are overcome if the proposed method is used. The input data for our algorithm are RGBE images. We need to convert from the RGBE image to the RGB image, and then convert from the RGB image to the HSV image. Luminance is obtained from the HSV image to perform the tone mapping process. Tone mapping consists of global tone mapping and local tone mapping. First, a global histogram adjustment method is applied for the initial global tone mapping. After that, a retinex-based local tone mapping is applied, which uses L_0 smoothing filter as the surround function to reduce halo artifacts. Finally, the result image is obtained from the tone-mapped luminance and the original RGB image through color correction. The following are the details of the proposed method.

3.1 Global Tone Mapping

The global tone mapping that is applied to luminance image performs a first compression of the dynamic range. It is like the early stage of the human visual system where a global adaptation takes place [7]. In the HDR image, since the luminance of bright area is too high, we firstly compress the high-luminance areas using this formula:

$$L(x, y) = \frac{L_w(x, y)}{1 + L_w(x, y)} \quad (5)$$

in which, $L_w(x, y)$ is the luminance of original image for pixel (x, y) , scaled to 0 and 100, while L is the initial compressed image.

With formula (5), the mapped image appears with low contrast, so it needs further processing. A method called histogram adjustment based linear to equalized quantizer (HALEQ) [8] is applied to enhance the contrast of image. Linear mapping divides luminance range into $N=256$ equal length intervals, and l_n is the cutting points. Histogram equalized mapping divides luminance range into N intervals such that the number of pixels falling into each interval is the same, and e_n is the cutting points. HALEQ strikes a balance between the linear mapping and the histogram equalized mapping, and the cutting points le_n is defined as this formula:

$$le_n = l_n + \beta(e_n - l_n) \quad (6)$$

in which, $0 \leq \beta \leq 1$ is parameter to control the global contrast. If $\beta = 0$, HALEQ is the linear mapping, $\beta = 1$, then HALEQ is the histogram equalized mapping. Pixels falling into the same interval are mapped to the same integer display level. We can express the whole process as

$$L_g(x, y) = \text{HALEQ}(L(x, y)) \quad (7)$$

In the above formula, L_g is the result of global tone mapping.

3.2 Local Tone Mapping

After the global adaptation processing, the local tone mapping based on retinex theory is applied to enhance the local contrast and preserve the details. The use of Gaussian function as surround function in retinex algorithm will cause the halo artifacts along high-contrast edges. But the halo artifacts can be reduced by introducing an edge-preserving filter, so L_0 smoothing filter is introduced as surround function in our algorithm. The local tone mapping can be expressed as the following formula:

$$L_{out}(x, y) = \text{LogF}(L_g(x, y)) - \text{LogF}(W(x, y)) \quad (8)$$

in which L_{out} is the result of local adaptation; W is the L_0 -smoothed version of L_g ; and $\text{LogF}(\cdot)$ represents a logarithm function as the follow formula:

$$\text{LogF}(L) = \frac{\log(100L + \varepsilon)}{\log(100 + \varepsilon)} \quad (8)$$

where ε refers to the nonlinearity offset. The logarithm function is a nonlinear function whose gradient is gradually decreasing. A small ε can effectively enhance the contrast of dark areas.

3.3 Color Correction

After the local adaptation, the processed luminance values are rescaled from 0 to 1. Finally, the tone mapped image is obtained from the processed luminance and the original RGB image. Similar to other approaches [7], the color correction is processed as the following formula:

$$\begin{bmatrix} R_{out} \\ G_{out} \\ B_{out} \end{bmatrix} = \begin{bmatrix} \left(\frac{R_{in}}{L_w} \right)^s \cdot L_{out} \\ \left(\frac{G_{in}}{L_w} \right)^s \cdot L_{out} \\ \left(\frac{B_{in}}{L_w} \right)^s \cdot L_{out} \end{bmatrix} \quad (8)$$

in which, R_{in} , G_{in} and B_{in} represent the original RGB image; R_{out} , G_{out} and B_{out} represent the tone mapped RGB image. L_w is the luminance of original image, and L_{out} is the luminance after local tone mapping. The exponent s controls the color saturation, which is usually set between 0.4 and 0.6. The result is then linearly scaled to the range of 0-255 for visualization.

4 Results

In this section, we have experimented, with our algorithm, on a variety of HDR images to compare with other three tone mapping operators. The three representative algorithms are the fast bilateral filter method [11], the photographic method [12], and the retinex-based adaptive filter method [13]. The parameters and the default values used for our experiments are shown in Table 1.

Table 1.

Parameters	Value
λ	0.0002
β	0.8
ε	0.001
s	0.4

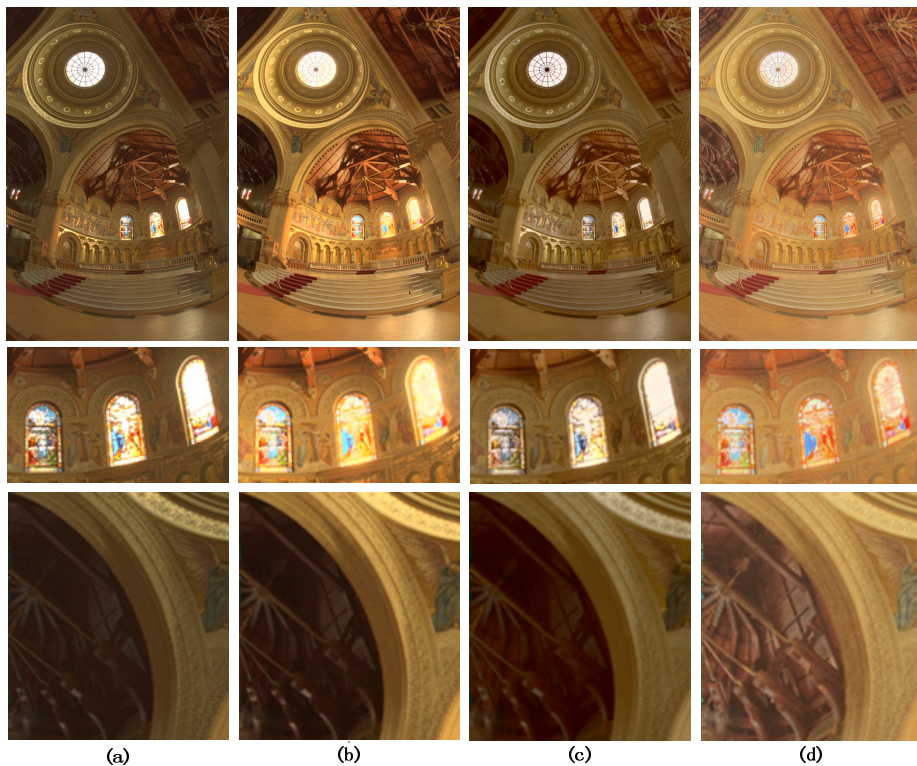


Fig. 2. Tone mapping results of the Memorial Church image. (a) Fast bilateral filter. (b) Photographic. (c) Retinex-based adaptive filter. (d) The proposed method.

The parameter λ controls the smoothness of image, and a small λ produces better dynamic range compression but reduces the global contrast. The parameter β controls the overall brightness of the image, and a large β makes the result image brighter. The parameter ε is the nonlinearity offset of logarithm function, and a small ε can effectively enhance the contrast of dark areas. The parameter s controls the color saturation of the image.

The comparison of our algorithm with other tone mapping operators is presented in Fig. 2. In order to make the comparison fairly, all the methods are tested with default parameters. Since there is no computational model to evaluate tone mapping algorithms objectively, we evaluate the results in a subjective way. Obviously, our algorithm shows a better performance in terms of naturalness and local contrast than other three algorithms. The middle row and the bottom row in Fig. 2 are respectively the enlarged images of bright and dark areas, and the result of our algorithm shows the clearest details of the four algorithms. In a word, our algorithm shows good color rendition while maintaining good contrasts elsewhere, and at the same time reduces halo artifacts.

5 Conclusion

In order to compress the dynamic range of the HDR images for display, a local tone mapping algorithm is proposed in this paper. In the global adaptation, we adopt HALEQ to enhance the global contrast. The local adaptation is based on retinex theory, and we use L_0 smoothing filter as surround function instead of Gaussian function to reduce the halo artifacts. The global and local tone mapping is only applied to luminance. Finally, the tone mapped image is obtained from the processed luminance and the original RGB image through color correction. The experimental results demonstrate that the proposed method effectively compress the dynamic range of image while with good color rendition and detail preservation.

However, since the proposed method uses L_0 smoothing filter in local tone mapping, it will be much slower compared with the global tone mapping algorithms. Future work is to optimize the computational model and also try to realize the parallel algorithm on GPU.

References

1. Larson, G.W., Rushmeier, H., Piatko, C.: A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes. *IEEE Transactions on Visualization and Computer Graphics* 3, 291–306 (1997)
2. Land, E., McCann, J.: Lightness and Retinex Theory. *J. Opt. Soc. Amer.* 61(1), 1–11 (1971)
3. Jobson, D.J., Rahman, Z., Woodell, G.A.: Properties and Performance of a Center/Surround Retinex. *IEEE Trans. Image Processing* 6(3), 451–462 (1997)
4. Xu, L., Lu, C., Xu, Y., Jia, J.: Image Smoothing via L_0 Gradient Minimization. *ACM Trans. Graph* 30, 6 (2011)

5. Paris, S., Durand, F.: A Fast Approximation of the Bilateral Filter Using a Signal Processing Approach. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. Part IV. LNCS, vol. 3954, pp. 568–580. Springer, Heidelberg (2006)
6. Farbman, Z., Fattal, R., Lischinski, D., Szeliski, R.: Edge-preserving Decompositions for Multi-scale Tone and Detail Manipulation. *ACM Trans. Graph.* 27, 3 (2008)
7. Alleysson, D., Süsstrunk, S.: On Adaptive Non-linearity for Color Discrimination and Chromatic Adaptation. In: Proc. IS&T First European Conf. Color in Graphics, Image, and Vision, Poitiers, France, April 2-5, pp. 190–195 (2002)
8. Duan, J., Bressan, M., Dance, C., Qiu, G.: Tone-mapping High Dynamic Range Images by Novel Histogram Adjustment. *Pattern Recognition* 43(5), 1847–1862 (2010)
9. Yun, B.J., Park, J., Kim, S., et al.: Color Correction for High Dynamic Range Images Using a Chromatic Adaptation Method. *Optical Review* 20(1), 65–73 (2013)
10. Mantiuk, R., Mantiuk, R., Tomaszewska, A., Heidrich, W.: Color Correction for Tone Mapping. *Computer Graphics Forum (Proc. Eurographics)* 28(2), 193–202 (2009)
11. Durand, F., Dorsey, J.: Fast Bilateral Filtering for the Display of High-dynamic-range Images. *ACM Transactions on Graphics (TOG)* 21(3), 257–266 (2002)
12. Reinhard, E., Stark, M., Shirley, P., Ferwerda, J.: Photographic Tone Reproduction for Digital Images. *ACM Trans. Graphics* 21(3), 267–276 (2002)
13. Meylan, L., Susstrunk, S.: High Dynamic Range Image Rendering With a Retinex-Based Adaptive Filter. *IEEE Trans. Image Processing* 15(9) (September 2006)

A Fast Mode Decision Algorithm and Its Hardware Design for H.264/AVC Intra Prediction

Wei Wang¹, Yuting Xie, Tao Lin, and Jie Hu

¹ College of Electronics Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
860290813@qq.com

Abstract. This paper presents an architecture of mode decision algorithm for H.264/AVC intra prediction. In the algorithm design, based on the inherent correlation existing in the spatial prediction modes, a significant computational savings can be achieved. In the hardware design, through efficient sharing of configurable units and parallel executions of different candidate prediction modes, a lower hardware utilization and higher execution speed can be achieved. Synthesis results show that the proposed architecture can process HDTV (1920×1080) video at 60 fps in FPGA platform and maximum frequency achieved is 184.8 MHz.

Keywords: H.264/AVC, intra prediction, spatial correlation, hardware verification and FPGA.

1 Introduction

With the wide use of technologies such as online video, digital television and video conference, video compression technology has become an inevitable component for storage and transmission. Currently, H.264/AVC is published as the new generate video compression coding standard which achieves the highest compression performance without sacrificing the quality of picture [1].

The high video compression efficiency of the H.264/AVC standard is achieved through a particular combination of a number of encoding tools but rather any single feature. Intra prediction algorithm is the important part of the encoder which occupies the most calculation and generates a prediction for a macroblock based on the spatial redundancy. However, the better compression efficiency comes with a high computation complexity which makes it different from meeting the demand of the real time [2].

The intra prediction algorithms with lower complexity and high-speed computing is important for the real-time requirements. Huang et al. [3] proposed an algorithm with better efficiency on mode decision, but it suffered from a higher computational complexity in determining the partitioning. Yu et al. [4] put forward a fast algorithm in determining the partitioning to reduce the high computational complexity, but their argument shows the insufficiency of efficiency on mode decision.

An efficient architecture for intra prediction is proposed based on the inherent correlation existing in the intra prediction modes which can reduce the candidate

prediction modes of luma blocks to less than half. With the hardware design, two metrics need to be better balanced for a higher throughput which represents the clock cycles and the critical path.

The essay is organized as follows: the implemented fast mode decision algorithm is proposed in section 2; the intra prediction hardware architecture is explained in Section 3; the synthesis results are presented in Section 4; and Section 5 presents the conclusion.

2 Implemented Fast Mode Decision Algorithm of Intra Prediction

This algorithm gives a fast way of calculating the best prediction mode of intra prediction, and which is proposed based on the inherent correlation between the intra prediction modes and blocks. In this way, the candidate prediction modes can reduce to less than half.

2.1 The Fast Decision of Partitioning in H.264 Intra Prediction

As mentioned above, the 4×4 intra prediction is preferred in the macroblock which has high texture complexity, otherwise, the 16×16 intra prediction is been chosen. Therefore, we can determine the partition by using the calculated texture complexity of a macroblock before intra prediction coding. The texture complexity of luma blocks on the directions of vertical, horizontal and oblique 135° are defined in the following way:

$$TC_{vertical} = \frac{1}{m(m-1)} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} |P_{i,j+1} - P_{i,j}| \cdot \quad (1)$$

$$TC_{horizontal} = \frac{1}{m(m-1)} \sum_{j=0}^{m-1} \sum_{i=0}^{m-1} |P_{i+1,j} - P_{i,j}| \cdot \quad (2)$$

$$TC_{angle} = \frac{1}{(m-1)(m-1)} \sum_{j=0}^{m-1} \sum_{i=0}^m |P_{i-1,j+1} - P_{i,j}| \cdot \quad (3)$$

where m is the size of luma block, $P_{i,j}$ is the original pixel values located in the i row and j column of the luma blocks.

The mean value of the texture complexity is computed as:

$$M_{TC} = \frac{1}{3} (TC_{vertical} + TC_{horizontal} + TC_{angle}) \cdot \quad (4)$$

The threshold of texture complexity is computed as:

$$TC_{th} = \frac{1}{M_{TC}} (|TC_{vertical} - M_{TC}| + |TC_{horizontal} - M_{TC}| + |TC_{angle} - M_{TC}|) \cdot \quad (5)$$

A MB has high texture complexity while $TC_{th} > 1$, and I4MB is selected as the prediction block size; otherwise, I16MB is chosen.

2.2 Mode Decision for Chroma 8×8 Block Intra Prediction

In the intra prediction mode decision algorithm, the conventional calculation times of Rate-Distortion-Optimization is $4 \times (9 \times 16 + 4) = 592$, the prediction mode of chroma block is used as the large exterior loop. By reducing the number of chroma candidate prediction modes, a significant computational saving can be achieved.

The texture complexity of chroma blocks on the main prediction directions are calculated by above equations (1)~(5). And m is the size of chroma blocks. Based on the texture complexity calculation of the 8×8 chroma blocks, the main candidate prediction modes can be determined as shown in table 1.

Table 1. Mode decision for Chroma 8×8 blocks

The Threshold of TC_{th}	The minimum value	Candidate Modes of Chroma 8×8 Block
$TC_{th} > 1$	TC _{vertical}	mode 2 (Vertical)
	TC _{horizontal}	mode 1 (Horizontal)
	TC _{angle}	mode 3 (Plane)
	TC _{vertical}	mode 2 ,0 (DC)
$TC_{th} < 1$	TC _{horizontal}	mode 1 ,0 (DC)
	TC _{angle}	mode 3 ,0 (DC)

2.3 Mode Decision for Luma 4×4 Block Intra Prediction

In the mode decision of intra prediction, a macroblock can be divided into one 16×16 block or sixteen 4×4 blocks or four 8×8 blocks, which are used for executing the mode decision. Therefore, the mode decision of 8×8 blocks are correlated with the 4×4 blocks, and the main candidate prediction modes of the 4×4 luma blocks can be determined depending on the prediction mode of 8×8 block as described in table 2.

Table 2. Intra prediction mode decision of luma 4×4 blocks

prediction mode of Chroma 8×8 block	Candidate modes of luma 4×4 macroblock
Mode 0	Main candidate modes 0, 1, 2, 4
Mode 1	Main candidate modes 1, 2, 6, 8
Mode 2	Main candidate modes 0, 2, 5, 7
Mode 3	Main candidate modes 0, 1, 2, 3

2.4 Mode Decision for Luma 16×16 Blocks

Due to the similarity of the prediction modes between luma 16×16 blocks and chroma 8×8 blocks, the 16×16 intra prediction mode decision are shown in table 3.

Table 3. Intra prediction mode decision of luma 16×16 blocks

prediction mode of Chroma 8×8 blocks	Main candidate modes of luma 16×16 block
Mode 0	Mode 2 (DC)
Mode 1	Mode 1 (Horizontal) + Mode 2 (DC)
Mode 2	Mode 0 (Vertical) + Mode 2 (DC)
Mode 3	Mode 3 (Plane) + Mode 2 (DC)

In order to cut down the number of the intra prediction modes and reduce high calculation burden, we choose the Simple Sum of Absolute Difference (SAD) calculation as the rate control model, which is used to determine the best intra prediction mode according to the principle of the minimum SAD value before coding.

$$SAD = \sum_{(x,y) \in MB_k} |original(x,y) - predict(x,y)| \quad (6)$$

While the position (x,y) represents the location of the luma pixels in the macro-block or sub-block, original(x,y) represents the original pixel value, predict(x,y) represents the prediction pixel value.

2.5 The Comparison between the New Algorithm and Original Algorithm of Intra Prediction

Combine with the original algorithms, a new mode decision algorithm for H.264 intra prediction is proposed. The computational complexity comparison of the two algorithms is shown below as table 4 gives.

Table 4. The computational complexity comparison of two algorithms

Modes	Chroma 8×8 intra prediction	luma 4×4 intra prediction	luma 16×16 intra prediction
New	1 or 2	2	1
Original	4	9	4

According to the above analysis, it is evident that the new mode decision algorithm can reduce the calculation times to $1 \times (16 \times 2) = 32$, $1 \times (16 \times 2 + 1) = 33$, $1 \times 1 = 1$, $2 \times (16 \times 2) = 64$, $2 \times (16 \times 2 + 1) = 66$ or $2 \times 1 = 2$. A significant operation efficiency improving can be achieved.

3 FPGA Design of H.264 Intra Prediction Hardware Architecture

From the above analysis, the hardware architecture of the proposed algorithm can be designed as shown in figure 1.

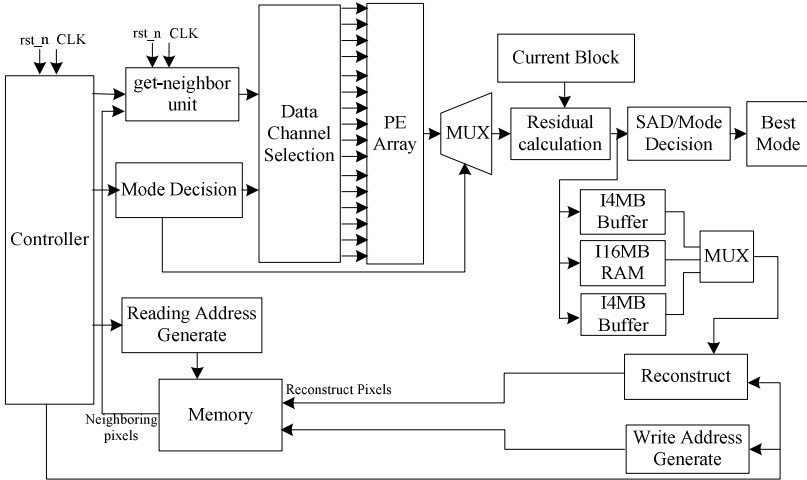


Fig. 1. Intra prediction hardware architecture

In the hardware design, several prediction modes are processed in parallel to generate the predict pixels for SAD calculation which is used for mode decision of intra prediction. After residual data generation and SAD calculation, the best prediction mode can be chosen, and the corresponding residual pixel value is input to the reconstruction loop for image reconstruction.

In summary, the architecture is composed by several modules: prediction generator, system control module, residual calculation module, SAD calculation module, SAD comparator module and mode decision module.

3.1 System Control Module

With the system control module, the best prediction mode and the partitioning of the current block can be determined. In addition, the scan sequence of the macroblock, the prediction sequence of the sub-blocks and the generation of reconstruction pixels are also generated by the system controller.

In the finite-state machine (FSM) design, the candidate prediction modes of intra prediction are specified in a certain order depending on the texture complexity value. When the best prediction mode of chroma 8×8 block is mode1, the FSM of luma 4×4 candidate prediction modes are described in figure 2.

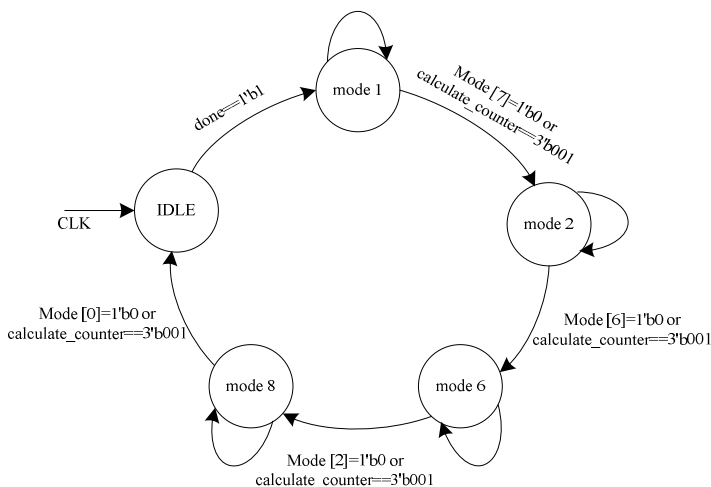


Fig. 2. State transition diagram of Control Unit

3.2 Predicted Generator Module

According to the fundamentals of intra prediction, it is proved that there are many common calculation parameters for different prediction modes, and many calculation formulas can be achieved in a same configurable calculation unit. In conclusion, we can obtain a configurable architecture for prediction generator module, and the processing element (PE) array is designed to generate concurrently 16 prediction pixels as shown in figure3.

(1) Parallel processing unit

Considering the clock frequency and the effective use of middle result registers, we proposed a parallel configurable and pipelining processing units to achieve the prediction calculation. Each PE generates one prediction pixel by selecting the right required reference pixel using multiplexers, and selects the right signal by a special logically controlled by FSM. However, the circuit will become a large scale, and requires a large capacity memory to match in the subsequent processing.

In the configurable architecture, each prediction element is composed of 3 components: the sum operation of the reference pixels, round value and shift value.

(2) Reconstructed neighbor samples memories and prediction memories

In intra prediction, the prediction processing for the current macroblock is existed when the reconstructed samples belonging to the neighbor blocks are available. For the real time processing of 60 fps of HD 1080p resolution, 486000 macroblocks should be processed per second, which result in high external memory bandwidth. This is why this work of research proposes a scheme, that a line of pixels in a frame is stored into the FPGA internal memory (BRAM).

For the luma 16×16 and Chroma 8×8 prediction mode decision is not executed in 4×4 block level, the re-processing result in a high number of clock cycles. In order to prevent the shortcomings, the architecture is taken into advantage of BRAMs in FPGA to store the predicted luma 16×16 blocks and Chroma 8×8 blocks.

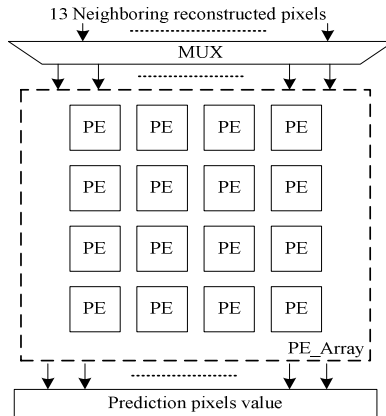


Fig. 3. Hardware architecture of PE array

4 Results and Comparisons with Related Work

The proposed algorithm is described in Verilog HDL and verified using Modelsim 6.5 SE. And then the Verilog RTL is synthesized to a Xilinx Virtex-5 FPGA using Synplify. The maximum clock frequency can be achieved at 184.8 MHz. In the simulation example, when the best prediction mode of chroma 8x8 block is mode1, the simulation waveform of prediction mode decision is shown in figure4. In the simulation waveform, the mincost port represents the minimum SAD value. The best mode port represents the best prediction mode corresponding to the minimum SAD value, which is mode 1 in this algorithm design.

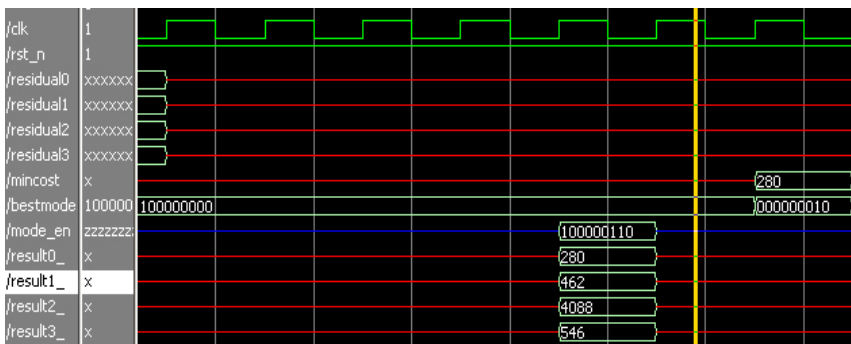


Fig. 4. The simulation waveform of intra prediction mode decision

The synthesis results of the architecture are shown in Table 5; and table 6 shows the comparisons with related work.

Table 5. Synthesis Result Of Intra Prediction

FPGA Device	Xilinx Virtex 5
Pixel Parallelism	16 pixels
36Kb BRAMs	10
Max.freq. (MHz)	184.8
LUTs	4699
Throughout (MB/s)	2.888 M

Table 6. Contrasts With Related Work

	[5]	[6]	[7]	This work
FPGA Device	Altera Stratix II	Altera Stratix II	Xilinx Virtex 2	Xilinx Virtex 5
Cycles/MB	36	--	--	<160
Max.freq/MHz	98.43	153	120	184.8
LUTs	3267	--	16546	4699

In the proposed architecture, with pipelining and parallel architecture, we can obtain 16 prediction pixels concurrently. Moreover, the throughput achieved by the proposed architecture is higher than the published results in [5,6,7], which make it possible to satisfy the requirement for the real time encoding.

5 Conclusion

An effective architecture is designed for intra prediction algorithm. By exploiting the inherent spatial correlation existing in the neighbor pixels and prediction modes, a significant computational saving can be achieved. With the hardware design, a parallel and configurable architecture is adopted to speed up the encoding time and at the same time it allows to reduce the computational complexity without any coding performance loss. The maximum clock frequency of the proposed hardware architecture can achieve 184.8 MHz in the Xilinx Virtex-5 FPGA. The experimental results confirm that the architecture can completely satisfy the real-time requirement for HDTV (1920×1080) video at 60 fps.

References

1. Wiegand, T., Sullivan, G.J., Bjøntegaard, G.: Overview of the H. 264/AVC Video Coding Standard. *Circuits and Systems for Video Technology* 13(7), 560–576 (2003)
2. Sahin, E., Hamzaoglu, I.: An Efficient Hardware Architecture for H.264 Intra Prediction Algorithm. In: *Design, Automation & Test in Europe Conference & Exhibition*, pp. 1–6. Nice (2007)

3. Huang, Y.H., Ou, T.S., Chen, H.H.: Fast Decision of Block Size, Prediction Mode, and Intra Block for H.264 Intra prediction. *Transactions on Circuits and Systems for Video Technology* 20(8), 1122–1132 (2010)
4. Yu, Y., Wang, L.: A fast Intra Mode Selection Method for H.264 High Profile. In: *International Conference on Acoustics, Speech and Signal Processing, Las Vegas*, pp. 681–684 (2008)
5. Palomino, D., Corrêa, G., Diniz, C., Bampi, S.: Algorithm and Hardware Design of a Fast Intra-frame Mode Decision Module for H.264/AVC encoders. In: *SBCCI 2011 Proceedings of the 24th Symposium on Integrated Circuits and Systems Design, New York*, pp. 143–148 (2011)
6. Shrivastava, V.K., Muralidhar, P., Rama Rao, C.B.: Architecture for H.264 Intra Prediction Fast Mode Decision Algorithm. *International Journal of Computer Applications* 68(7), 1–6 (2013)
7. Li, X.Y., Ji, F.: A Parallel H.264 Intra-Frame Prediction Decision Architecture Based on FPGA. In: *International Conference on Computational and Information Sciences (ICCIS), Shiyang*, pp. 1611–1615 (2013)

Research of Multi-focus Image Fusion Algorithm Based on Sparse Representation and Orthogonal Matching Pursuit

Li Xuejun^{1,2} and Wang Minghui¹

¹ Sichuan University, College of Computer Science, Chengdu, 610065

² Southwest University of Science and Technology, Mianyang, 621010, China

lixuejunmai@163.com, wangminghui@scu.edu.cn

Abstract. Due to the unideal effects of those common multi-source focus image fusion algorithms, in this essay we propose a multi-focus image fusion algorithm based on sparse representation and orthogonal matching pursuit (OMP), and demonstrate the results of the corresponding multi-source focus image fusion experiments by MATLAB. Compared with the fused images of the above several common algorithms by evaluating subjectively and objectively, the results suggest that the multi-focus image fusion algorithm based on sparse representation and orthogonal matching pursuit (OMP) present higher mutual information, minimum distorted values and higher $Q^{ab/f}$ values which indicate that the fused image by this algorithm can obtain more image information with a smaller distortion from the original (image?), so as to get a better image but cost much more time.

Keywords: Multi-focus image fusion, sparse representation, orthogonal matching pursuit and performance evaluation.

1 Introduction

The clarity of optical lens imaging relies on its depth of field in an optical imaging system. If the object distance of a target in the scene goes beyond the depth of field to the optical lens, the image will be fuzzy, whereas a clear one can be assured. In reality, the targets in the same scene produce different clarities on the impact of external factors such as the distance and the light intensity when imaging. Therefore, it is quite difficult to make an accurate and comprehensive interpretation from the information of a single image in the same scene. Multi-focus image fusion technique is used to deal with two or more images (abandoning the fuzzy parts and keeping the clear parts), which are shot by different focus objects with different depths of field in the same scene. It will finally form a new image to be perceived more suitably and processed more easily. Multi-focus image fusion is a research branch of multi-source image fusion [1].

Suppose that there are K images $\{I_i\}_{i=1}^K$ which describe the same scene with subjects in different depths of field, but focus on different objects respectively, the problem to be solved by multi-focus image fusion is how to recover the image F that

all focus objects are clear from those K images. If $\mathcal{F}(\cdot)$ represents the fusion operator, the multi-focus image fusion can be described by the following formula:

$$F = \mathcal{F}(I_1, I_2, \dots, I_k) \quad (1)$$

As to the multi-focus image, the classic multi-source image fusion algorithms (based on FSD pyramid image fusion algorithm[2-4], DWT image fusion algorithm[5], the contrast pyramid image fusion algorithm, SIDWT image fusion algorithm[6-7] and the spatial frequency fusion algorithm[8-11]) are still adopted to deal with the corresponding image fusion which can be evaluated subjectively and objectively. Five common objective evaluation indicators include mutual information(MI), average gradient(AG), correlation coefficient(CC), distorted degree(DD) and $Q^{ab/f}$. The experimental results suggest that the effects of those common multi-source focus image fusion algorithms are not ideal, this is why, in this essay, we propose a multi-focus image fusion algorithm based on sparse representation and orthogonal matching pursuit.

2 Multi-focus Image Fusion Algorithm Based on Sparse Representation

2.1 Image Sparse Representation and Orthogonal Matching Pursuit Algorithm

The core of sparse representation is that signals can be approximated by the linear combination of a small number of columns in the over-complete dictionary D (each column in D is also called an atom). But how to obtain the coefficient of the linear combination becomes the main problem of sparse representation, and the following formula can be described:

$$\min \|x\|_0, \text{subject to } \|y - Dx\|_2^2 \leq \epsilon \quad (2)$$

In formula 2, $\|x\|_0$ represents l_0 to calculate numbers of non-zero components in vector x , and ϵ is allowable approximation error.

However, solving formula 2 to make x satisfy $y = \Phi x$ is a NP-hard problem without being solved accurately in computer time. But thanks to the efforts of scientists, the above problem resolution has been transformed from the unsolvable l_0 minimization to a minimizing model of solving l_1 by some convex optimization algorithms, such as interior point method[12], gradient projection method[13-14], greedy algorithm[15-16], iterative threshold method[17-18] and Bregman iteration based on Bregman distance[19]. Recently, the iterative greedy algorithm has attracted much attention by its low complexity and simple geometric explanation, aiming to obtain reconstruction by the support of iteration to calculate x . It mainly includes matching pursuit: MP [20], orthogonal matching pursuit: OMP[21], stagewise orthogonal matching pursuit: StOMP[22], regularized orthogonal matching pursuit: ROMP[23], compressive sampling matching pursuit: CoSaMP[24], subspace pursuit: SP[25] and improved backward optimized OMP: IBOOMP [26], etc.

This essay raises orthogonal matching pursuit algorithm to solve formula 2 and then to realize the reconstruction of the signal. Orthogonal matching pursuit algorithm is the improvement of matching pursuit algorithm which is an intuitive iterative greedy algorithm for the minimizing model to solve, whose basic idea is to give a more complete dictionary database and make the signal approach gradually shown. During the process of each approach, orthogonal matching pursuit algorithm screens atoms through orthogonal matching. It is a greedy algorithm that tends to find out the most relevant atom to the residual signal of D in each iteration, solve the corresponding value of x , and finally update the residual signal. The algorithm flowchart is as shown in figure 1:

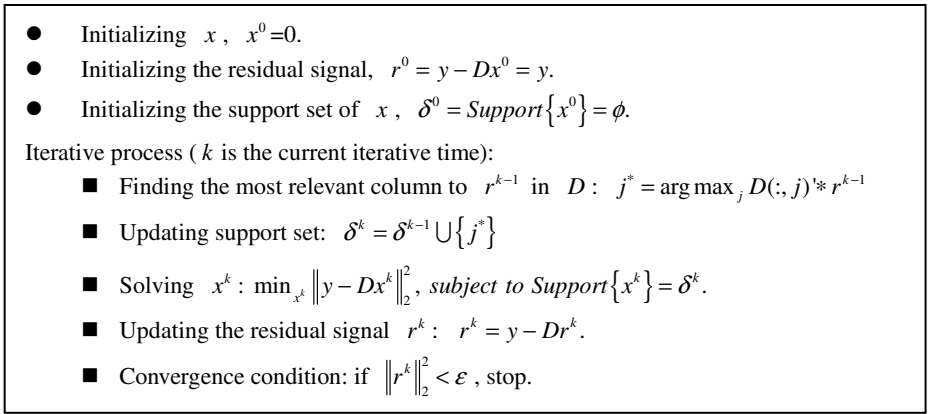


Fig. 1. OMP Algorithm Flowchart

One can obtain the over-complete dictionary in the algorithm by using over-complete DCT dictionary or K-SVD dictionary learning algorithm. The K-SVD dictionary learning algorithm is based on singular value decomposition for sparse representation proposed by Mallat in 2006. Its algorithm is as the followings: firstly initialize a dictionary D , and solve the sparse coefficient X by sparse representation algorithm; then remain the position of non-zero values in solved sparse coefficient X , to calculate the new D and X by SVD (singular value decomposition). The following mathematical formula can be described:

$$\operatorname{argmin}_{D,X} \left(\sum_i \|X(:,i)\|_0 + \lambda \|D - DX\|_F^2 \right), \text{subject to } \|D(:,i)\|_2^2 = 1 \quad (3)$$

2.2 Algorithm Implementation

Image Sparse Representation. In the specific process of algorithm implementation, one needs to divide an image of $W \times H$ into overlapping image blocks of $n \times n$ when using sparse representation for image processing. That means that the image can be overlapped with the upper-left corner of a small window to be $n \times n$ in itself when taking the first image block; then the small window moves one pixel to the right, and the second image block can be received. Until it can't move to the right, the small

winder goes back to the leftmost position and then moves down for one pixel. And so on, an image of $W \times H$ can be divided into $(W + n - 1) \times (H + n - 1)$ image blocks of $n \times n$. The advantages of using overlapping image blocks are:

1) Images in any size can be divided into blocks without taking boundary treatment into consideration;

2) It can avoid blocking effects to some extent when taking image reconstruction.

Suppose that there are K image: $\{I_i\}_{i=1}^K$, and each image is divided into overlapping image blocks of $n \times n$ to get $\{P_i\}_{i=1}^K$, in which every column of P_i is obtained by the form of column vector from an image block of $n \times n$. If I_i is the image of $W \times H$, P_i is the matrix of $\{(n \times n)\} \times \{(W + n - 1) \times (H + n - 1)\}$. The m -column $P_i(:, m)$ of P_i can be got its corresponding sparse coefficient $X_i(:, m)$ by sparse representation based on over-complete dictionary D . If solving the sparse coefficient of $\{P_i\}_{i=1}^K$, we can get K matrixes $\{X_i\}_{i=1}^K$.

Coefficients Fusion. Suppose that it exists an image F , with dividing it into blocks to get $P^{(F)}$ firstly. And then the sparse coefficient $X^{(F)}$ can be obtained by sparse representation. If $X^{(F)}$ is known, the image F can appear through the inverse process.

This algorithm uses solved $\{X_i\}_{i=1}^K$ to get $X^{(F)}$. The m -column $X^{(F)}(:, m)$ of $X^{(F)}$ is replaced by the maximum l_1 norm in the m -column of each matrix in $\{X_i\}_{i=1}^K$. And the following formulas can be expressed:

$$\begin{aligned} idx &= \arg \max_i \|X_i(:, m)\|_1 \\ X^{(F)}(:, m) &= X_{idx}(:, m) \end{aligned} \quad (4)$$

The fusion strategy is used to recover per column of $X^{(F)}$ and finally the image F can be reconstructed by the inverse process.

3 Image Fusion Experiments and Results

3.1 Image Fusion Rules

In the process of the corresponding multi-source image fusion algorithm implementation, we select four groups of multi-focus test images in this essay, with two images for each group. For multi-focus images in each group, five classic multi-source image fusion algorithms (based on FSD Laplacian pyramid transform, Contrast pyramid transform, DWT with DBSS (2,2) wavelet transform, SIDWT with Haar translational invariance discrete wavelet transform and spatial frequency fusion algorithm), sparse representation and orthogonal matching pursuit multi-focus image fusion algorithm proposed in the paper are applied to do image fusion experiments by MALAB. To make fused images based on various multi-source fusion algorithms

comparable, unified rules are adopted in terms of integration parameters selection. In the decomposition of pyramid, high-frequency coefficients take the maximum and the low-frequency coefficients take the average. When the images are in transformation, layers are all taken as 4. In the spatial frequency fusion algorithm, different types of test images lead to different sizes of the best Block. According to subjective effects of fused images by experiments, in this essay, we select the best Block size of multiple images to be 8×8 , and the threshold TH to be 1. The dictionary of multi-focus image fusion algorithms based on sparse representation and orthogonal matching pursuit has used over-complete DCT dictionary and K-SVD dictionary, and the size of the image block is 8×8 .

3.2 Image Fusion Results and Subjective and Objective Evaluations

(1) Subjective and objective evaluations of each image fusion algorithm to the multi-focus image of group A

Original multi-focus image data and fusion results of group A are shown in figure 2, and objective evaluation indicator data are shown in table 1.

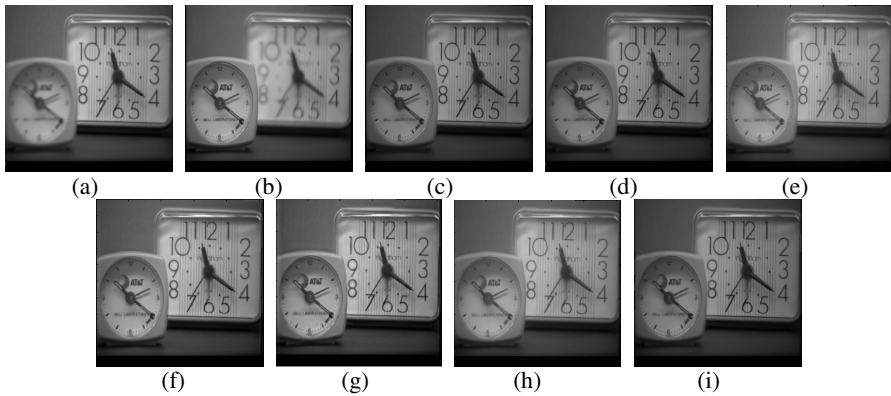


Fig. 2. Experimental Results of Multi-focus Image Fusion of Group A. (a)Foreground Fuzzy Image (b)Background Fuzzy Image (c) 8×8 DCT (d) 8×8 K-SVD (e)FSD Algorithm (f)Contrast Algorithm (g)DWT Algorithm (h)SIDWT Algorithm (i)Spatial Frequency Algorithm.

Table 1. Objective Evaluation Indicator Data to Multi-focus Fusion Image of Group A

Algorithm	MI	AG_A	AG_B	AG_F	CC_A	CC_B	DD_A	DD_B	$Q^{AB/F}$
8×8 DCT	7.903	2.550	1.827	2.790	0.991	0.983	0.946	2.332	0.698
8×8 KSVD	7.467	2.550	1.827	2.662	0.991	0.983	1.007	2.125	0.683
FSD	6.231	2.550	1.827	2.830	0.989	0.989	2.124	2.172	0.663
Contrast	6.873	2.550	1.827	3.747	0.993	0.982	2.159	2.199	0.630
DWT	6.172	2.550	1.827	3.847	0.989	0.979	2.629	2.243	0.604
SIDWT	6.627	2.550	1.827	3.661	0.992	0.982	2.675	2.262	0.671
SF	8.058	2.550	1.827	3.452	0.993	0.983	2.225	2.212	0.685

(2) Subjective and objective evaluations of each image fusion algorithm to the multi-focus image of group B

Original multi-focus image data and fusion results of group B are shown in figure 3, and objective evaluation indicator data are shown in table 2.

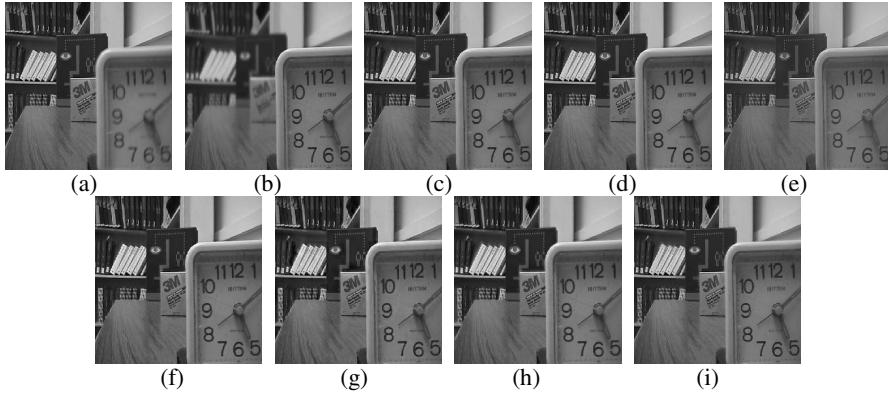


Fig. 3. Experimental Results of Multi-focus Image Fusion of Group B. (a)Foreground Fuzzy Image (b)Background Fuzzy Image (c)8*8 DCT (d)8*8 K-SVD (e)FSD Algorithm (f)Contrast Algorithm (g)DWT Algorithm (h)SIDWT Algorithm (i)Spatial Frequency Algorithm.

Table 2. Objective Evaluation Indicator Data to Multi-focus Fusion Image of Group B

Algorithm	MI	AG_A	AG_B	AG_F	CC_A	CC_B	DD_A	DD_B	$Q^{AB/F}$
8*8DCT	7.362	4.789	6.896	7.611	0.967	0.985	4.024	1.123	0.723
8*8KSVD	7.244	4.789	6.896	7.602	0.967	0.985	3.892	1.052	0.723
FSD	5.217	4.789	6.896	6.376	0.967	0.980	4.698	3.372	0.669
Contrast	6.270	4.789	6.896	7.924	0.969	0.985	4.466	1.908	0.689
DWT	5.612	4.789	6.896	8.092	0.967	0.984	4.638	2.264	0.667
SIDWT	5.948	4.789	6.896	7.585	0.972	0.986	4.302	2.073	0.698
SF	7.437	4.789	6.896	7.456	0.968	0.985	3.788	1.073	0.710

(3) Subjective and objective evaluations of each image fusion algorithm to the multi-focus image of group C

Original multi-focus image data and fusion results of group C are shown in figure 4, and objective evaluation indicator data are shown in table 3.

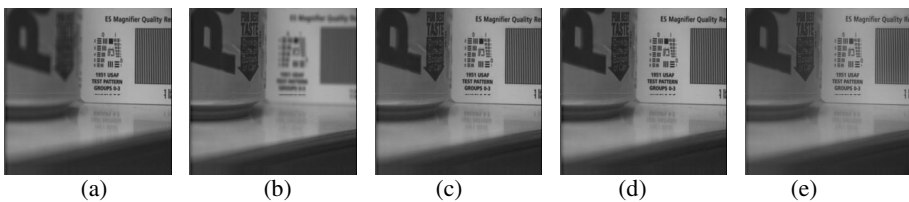


Fig. 4. Experimental Results of Multi-focus Image Fusion of Group C. (a)Foreground Fuzzy Image (b)Background Fuzzy Image (c)8*8 DCT (d)8*8 K-SVD (e)FSD Algorithm (f)Contrast Algorithm (g)DWT Algorithm (h)SIDWT Algorithm (i)Spatial Frequency Algorithm.

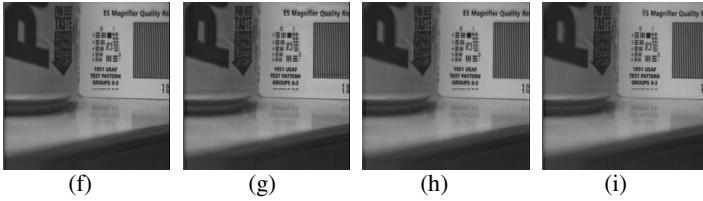


Fig. 4. (Continued)

Table 3. Objective Evaluation Indicator Data to Multi-focus Fusion Image of Group C

Algorithm	MI	AG_A	AG_B	AG_F	CC_A	CC_B	DD_A	DD_B	Q^{ABF}
8*8DCT	7.923	4.026	2.753	4.323	0.998	0.974	0.935	2.780	0.753
8*8KSVD	7.792	4.026	2.753	4.302	0.998	0.974	0.880	2.712	0.753
FSD	5.96	4.026	2.753	3.646	0.994	0.971	2.496	3.616	0.713
Contrast	7.161	4.026	2.753	4.427	0.997	0.974	1.278	2.687	0.741
DWT	6.447	4.026	2.753	4.562	0.996	0.975	1.481	2.805	0.712
SIDWT	6.833	4.026	2.753	4.201	0.996	0.947	1.455	2.946	0.728
SF	7.814	4.026	2.753	4.259	0.998	0.974	0.957	2.643	0.747

(4) Subjective and objective evaluations of each image fusion algorithm to the multi-focus image of group D

Original multi-focus image data and fusion results of group D are shown in figure 5, and objective evaluation indicator data are shown in table 4.

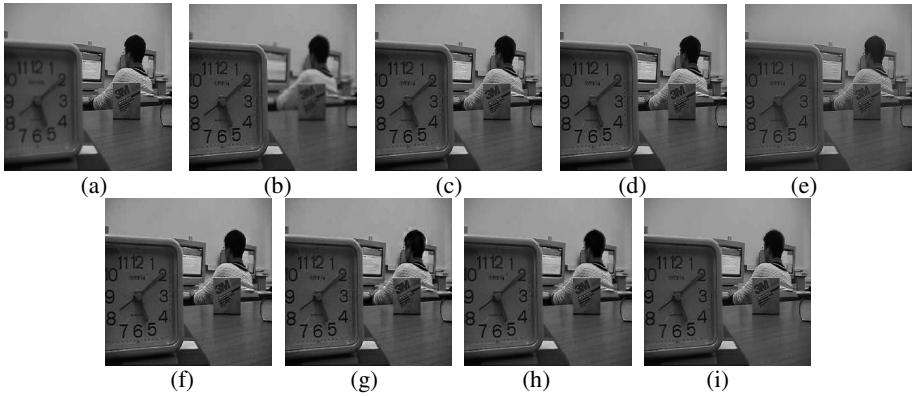


Fig. 5. Experimental Results of Multi-focus Image Fusion of Group D. (a)Foreground Fuzzy Image (b)Background Fuzzy Image (c)8*8 DCT (d)8*8 K-SVD (e)FSD Algorithm (f)Contrast Algorithm (g)DWT Algorithm (h)SIDWT Algorithm (i)Spatial Frequency Algorithm

Table 4. Objective Evaluation Indicator Data to Multi-focus Fusion Image of Group D

Algorithm	MI	AG_A	AG_B	AG_F	CC_A	CC_B	DD_A	DD_B	$Q^{ab/f}$
8*8DCT	7.843	3.912	4.402	5.225	0.997	0.996	2.827	0.983	0.731
8*8KSVD	7.582	3.912	4.402	5.178	0.976	0.996	2.702	0.975	0.727
FSD	5.640	3.912	4.402	4.428	0.979	0.988	3.702	2.779	0.675
Contrast	7.059	3.912	4.402	5.472	0.981	0.993	2.919	1.566	0.691
DWT	6.454	3.912	4.402	5.616	0.981	0.991	2.904	1.710	0.666
SIDWT	6.789	3.912	4.402	5.257	0.984	0.992	2.750	1.544	0.695
SF	8.049	3.912	4.402	5.135	0.977	0.996	2.815	0.910	0.732

4 Conclusion

Through the careful observation of four groups of multi-focus fusion image by Matlab and comparing with the corresponding data for objective evaluation index, we can draw the following conclusions: from the view of fusion image, the results of image fused by various algorithms differ slightly. But from the details of clarity and texture, the fused images by the multi-focus image fusion algorithm based on sparse representation and orthogonal matching pursuit have better clarity and texture details, with more information from the original what?. From the objective evaluation index, the multi-focus image fusion algorithm based on sparse representation and orthogonal matching pursuit presents higher mutual information, larger average gradient value AG, minimum distortion coefficients DD and higher $Q^{ab/f}$ values, which means that this image fusion algorithm can maintain more original information with the smallest distortion, reflecting the edge information and the importance of the original image. Therefore, it makes better effects of image fusion than other multi-source image fusion algorithms.

References

1. Pohl, C., Van Genderen, J.L.: Multisensor Image Fusion in Remote Sensing: Concepts, Methods and Applications. *International Journal of Remote Sensing* 19(5), 823–854 (1998)
2. Eichmann, G.: Pyramidal Image Processing Using Morphology. *Applications of Digital Image Processing XI. Proceedings of the SPIE* 974, 30–37 (1998)
3. Goutsias, J., Heijmans, H.M.: Nonlinear Multi-resolution Signal Decompositions Scheme-Part1: Morphological Pyramids. *IEEE Trans. on Image Processing* 9(11), 1862–1876 (2000)
4. Burt, P.J., Kolczynski, R.J.: Enhanced Image Capture Through Fusion. In: *Fourth International Conference on Computer Vision*, pp. 173–183. IEEE Press, Berlin (1993)
5. Qiguang, M., Baoshu, W.: Multi-Sensor Image Fusion Based on Improved Laplacian Pyramid Transform. *Acta Optica Sinica* 29(9), 1605–1610 (2007)

6. Rockinger: Image Sequence Fusion Using a Shift-Invariant Wavelet Transform. In: International Conference on Image Processing (ICIP 1997), vol. III, pp. 288–292. IEEE Press, Washington, DC (1997)
7. Yu, L.-S., Wen, G.-J., Li, Z.-Y.: Remote Sensing Image Fusion Algorithm Based on Shift Invariance Discrete Wavelet Transform. *Computer Engineering* 37(17), 197–199 (2011)
8. Li, S., Wang, Y., Zhang, C.: Feature of Human Vision System Based Multi-Focus Image Fusion. *Acta Electronica Sinica* 29(12), 1699–1701 (2001)
9. Do, M.N., Martin, V.M.: The contourlet transform: An efficient directional multi-resolution image presentation. *IEEE Trans. on Image Processing* 14(12), 2091–2016 (2005)
10. Ma, X.-X., Peng, L., Xu, H.: Block-based Assimilation of Spatial Frequency Multi-focus Image Fusion Algorithm. *Science Technology and Engineering* 12(1), 64–67 (2012)
11. Miao, Q.J., Wang, B.S.: A Novel Image Fusion Method Using Contourlet Transform. In: Proceeding of 4th International Conference on Communications, Circuits and Systems (ICCCAS 2006), vol. 1, pp. 549–552. IEEE Press, Guilin (2006)
12. Candes, E., Tao, T.: Decoding by linear programming. *IEEE Transactions on Information Theory* 51(12), 4203–4215 (2005)
13. Candes, E., Recht, B.: Exact matrix completion via convex optimization. *Commun. ACM* 55(6), 111–119 (2012)
14. Candes, E.: Compressive sampling. *International Congress of Mathematicians* 52(4), 1289–1306 (2006)
15. Candes, E., Tao, T.: Near-optimal signal recovery from random projections: Universal encoding strategies. *IEEE Transactions on Information Theory* 52(12), 5406–5425 (2006)
16. Mendelson, S., Pajor, A., Jaegermann, N.T.: Uniform uncertainty principle for Bernoulli and subgaussian ensembles. *Constructive Approximation* 28(3), 277–289 (2008)
17. Bajwa, W.U., Haupt, J.D., Raz, M.G., Wright, S.J., Nowak, R.D.: Toilets-Structured Compressed Sensing Matrices. In: *IEEE/SP 14th Workshop on Statistical Signal Processing*, pp. 294–298. IEEE Press, Madison (2007)
18. Rauhut, H.: Compressive sensing and structured random matrices. *Theoretical Foundations and Numerical Methods for Sparse Recovery* 9, 1–92 (2010)
19. Donoho, D.L.: For most large underdetermined systems of linear equations, the minimal L1 norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics* 59(6), 797–829 (2006)
20. Mallat, S., Zhang, Z.: Matching pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing* 41(12), 3397–3415 (1993)
21. Tropp, J., Gilbert, A.: Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory* 53(12), 4655–4666 (2007)
22. Donoho, D.L., Tsaig, Y., Drori, I., Starck, J.L.: Sparse solution of underdetermined linear equations by stage wise orthogonal matching pursuit. Technical Report No. 2006-2, Department of Statistics, Stanford University, USA (2006)
23. Needell, D., Vershynin, R.: Signal recovery from inaccurate and incomplete measurements via regularized orthogonal matching pursuit. *IEEE Journal of Selected Topics in Signal Processing* 4(2), 310–316 (2010)
24. Needell, D., Tropp, J.A.: CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis* 26(3), 301–321 (2009)

25. Dai, W., Milenkovic, O.: Subspace pursuit for compressive sensing signal reconstruction. *IEEE Transactions on Information Theory* 55(5), 2230–2249 (2009)
26. Fang, H., Zhang, Q.B., Wei, S.: Image reconstruction based on improved backward optimized orthogonal matching pursuit algorithm. *Journal of South China University of Technology (Natural Science)* 36(8), 23–27 (2008)

Infrared Face Recognition Based on DCT and Partial Least Squares

Zhijhua Xie and Guodong Liu

Key Lab of Optic-Electronic and Communication,
Jiangxi Sciences and Technology Normal University, Nanchang, Jiangxi, 330013, China
xie_zhijhua68@aliyun.com

Abstract. Infrared face imaging, being light-independent, and not vulnerable to facial skin expressions and posture, can avoid or limit the drawbacks of face recognition in visible light. However, to obtain the compact and discriminative feature extracted from infrared face image is a challenging task. In this essay, infrared face recognition method using Discrete Cosine Transform (DCT) and Partial Least Square (PLS) is proposed. Due to strong ability for data decorrelation and compact energy, DCT is studied to obtain the compact features in infrared face. To make full use of the discriminative information in DCT coefficients, the final classifier formulates PLS regression for accurate classification. The experimental results show that the proposed algorithm outperforms Principle Component Analysis (PCA) and DCT based infrared face recognition algorithms.

Keywords: Infrared face recognition, Partial least square, feature extraction, discrete cosine transform.

1 Introduction

As we know, the resolution of the infrared image is lower than the visible image. This is to say that the infrared image has little local discriminative information.

For this reason, the compact and discriminative feature extraction from the infrared face image is a challenging task [4, 8]. In previous research, Discrete Cosine Transform (DCT) based on the feature extraction method is applied to extract compact information for face recognition [5]: Zhang et al [7] improved the classical face features extraction method (Principle Component Analysis (PCA) + Linear Discriminant Analysis (LDA)) and proposed DCT and LDA based on features extraction algorithm; Yin et al [6] improved DCT and LDA face recognition method using Feature Selection (FS) in DCT domain. As for infrared face recognition, Xie et al [8] applied DCT and FS to find a compact features extraction method. However, the discriminative performance of DCT features received less attention. The main idea in this essay is that different DCT coefficients do have different ability to discriminate various classes. In other words, some coefficients, namely discriminant coefficients, should have bigger weights than others. Therefore, how to make full use

of the discriminative information in DCT coefficients is a key step in order to obtain good performance of DCT based infrared face recognition method [5].

Partial Least Squares (PLS) is a supervised effective discriminative dimension reduction technique [11, 12] and has been successfully applied to many vision applications including face recognition [2, 9, 10]. In this essay, we use PLS to find a much smaller number of discriminative factors in DCT features. Experimental results show that PLS further improves the recognition performance based on DCT+LDA features. This is because PLS basis projects the feature vectors into a latent space in which feature vectors corresponding to the same subject are closer than the feature vectors corresponding to different subjects.

2 Discrete Cosine Transformation

The discrete cosine transformation (DCT) is a popular image compression method [5]. The nuclear transformation of the discrete cosine transformation is the cosine function of real, thus the calculation complexity of DCT is simple and its information packing ability closely approaches PCA. Another merit of the DCT is that it can be implemented efficiently using the Fast Fourier Transform (FFT).

For a $M \times N$ digital image $f(x, y)$, its two-dimensional DCT, $C(u, v)$ is shown in the following equation:

$$C(u, v) = a(u)a(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \times \cos \left[\frac{(2x+1)u\pi}{2M} \right] \cos \left[\frac{(2y+1)v\pi}{2N} \right] \quad (1)$$

$$u = 0, 1, \dots, M-1; v = 0, 1, \dots, N-1$$

where $C(u, v)$ is the result of DCT which is the DCT coefficient that represents the purpose of study in this essay. Please be aware that $a(u)$ 、 $a(v)$ are defined respectively as:

$$a(u) = \begin{cases} \sqrt{1/M} , & u = 0 \\ \sqrt{2/M} , & u = 1, 2, 3, \dots, M-1 \end{cases} \quad (2)$$

$$a(v) = \begin{cases} \sqrt{1/N} , & v = 0 \\ \sqrt{2/N} , & v = 1, 2, 3, \dots, N-1 \end{cases}$$

Based on high-compression characteristics and valuable information packing ability of DCT, it can be used for feature extraction of infrared face recognition to reduce the relevance of infrared face data [6, 7]. When reconstructing the image using DCT coefficient, retaining few low-frequency component of DCT and rounding down mostly high-frequency component, still get the restore images that is similar to the original images using anti-transformation. The original infrared face, corresponding to the DCT coefficients and the reconstructed image using 1 / 25 of the DCT coefficients are shown in Figure1. As we can see, the majority of important figure (including nose, mouth, cheeks, etc.) in the restore infrared face is preserved.

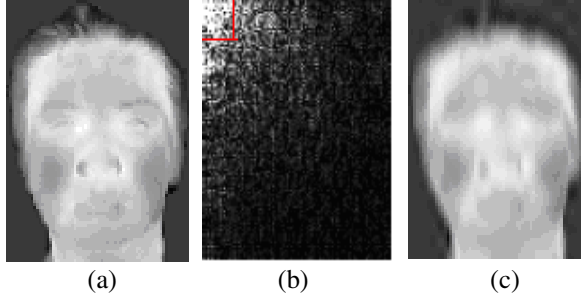


Fig. 1. The original map, corresponding to the DCT coefficients and the reconstructed image using 1 / 25 of the DCT coefficients. (a) is the normalized image. (b) is DCT coefficients (c) is the image using 1 / 25 of the DCT coefficients.

Theoretically, the number of DCT transformation coefficient equals the size of the image, as the previous analysis indicated, it can be observed that a large amount of information about the original image is stored in a fairly small number of coefficients (in the upper-left corner, corresponding to the low spatial frequency components in the image). Thus it is reasonable to take a certain amount of the DCT coefficient as a feature vector, and preserve it in the database. In this research paper, a form of rectangular window is used in order to preserve the DCT coefficients.

3 Partial Least Squares Classifier

In this essay, we extract and classify the features from the DCT transformation based on Partial Least Squares (PLS) regression. PLS is a supervised learning method that models linear relations between sets of independent variables and response variables via an intermediate latent space [12]. PLS models the relations between sets of observed variables by means of latent variables. In its general form, PLS creates orthogonal score vectors by maximizing the covariance between different variable sets [11, 14].

Let us suppose that c is the number of subject classes contained in the training database and $G_f = \{f_{ij}\}_{j=1,k=1}^{n_k,c} \in \mathfrak{R}^{d \times g}$ is the final gallery representation such

that $n_k \geq 1$ are the number of samples in each class, $g = \sum_{k=1}^c n_k$ f_{ij} is a DCT based

feature vector representing the j th sample in the k th person. Then, the training samples G_f denote a set of predictor variables. We represent the response variables

$Z = \{z_{ij}\}_{j=1,k=1}^{n_k,c}$ as a set of indicator vectors. Where $z_{ij} \in \mathfrak{R}^c$ shows the membership of k th class. z_{ij} is defined as a binary vector having 1 at the k th index and zeros otherwise. PLS decomposes matrices G_f and Z into the form [14].

$$G_f = BP^T + E \quad (3)$$

$$Z = UQ^T + F \quad (4)$$

Where B and U are the matrices containing the extracted latent vectors, the matrices P and Q represent loadings, and the matrices E and F are the residuals. Based on the nonlinear iterative partial least squares (NIPALS) algorithm [9] for learning the latent space, PLS finds weight vectors p and q such that

$$\text{cov}(b, u)^2 = \max_{p=q=1} \text{cov}(G_{fp}, Z_q)^2 \quad (5)$$

Where b and u are the column vectors of B and U respectively and $\text{cov}(b, u)$ is the sample covariance. The regression coefficients between the two sets of variables G_f and Z can be estimated by PLS regression formulation [10]

$$W = G_f^T U^T (B^T G_f G_f^T)^{-1} B^T Z \quad (6)$$

Using W , we can predict labels of the query feature vector f_t

$$\hat{z}_t = f_t^T W \quad (7)$$

Where $\hat{z}_t \mathfrak{K}^c$ is an indicator variable, ideally containing 1 at only one location (indicating the class membership) and 0 at all other locations. However \hat{z}_t contains some non-zero value at each location due to the noise in the data and approximation errors in the regression process. The location of the maximum of \hat{z}_t is considered as the predicted label for f_t .

Using this method, for each test sample, we can obtain c regression values \hat{z}_t from all the PLS classifiers. The category corresponding to the maximum value of \hat{z}_t is decided to be the recognition result.

4 Experimental Results

The infrared data in this essay were collected by using an infrared camera Thermo Vision A40 supplied by FLIR Systems Inc [4]. The training database comprises 500 thermal images of 50 individuals which were carefully collected under the similar conditions in November 17, 2006: environment under air-conditioned control with temperature around 25.6~26.3°C. The test database comprises 500 thermal images of 50 individuals were obtained under the same conditions of the training database. The original resolution of each image is 240×320. In our experiments, the face image is normalized to the size of 80×60.



Fig. 2. Part of infrared face database

The recognition rate of proposed infrared face recognition method is shown in Figure 3. It is evident on the Figure 3 that our proposed infrared face recognition method can reach the highest recognition rate (95.8%) when the number of PLS bases is 50. The PLS can be used to effectively reduce and discriminate the DCT coefficients in infrared face images.

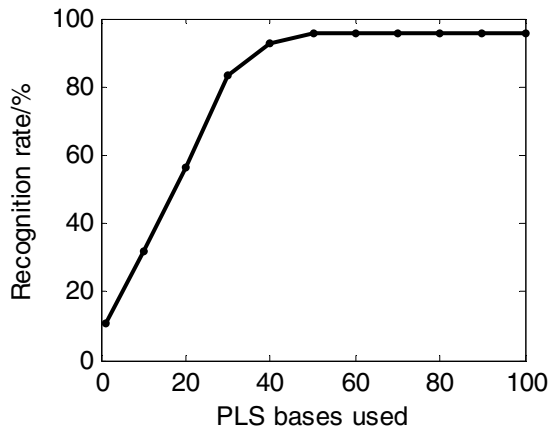


Fig. 3. Recognition rate vs the number of PLS bases used

To verify the effectiveness of the proposed features extraction method for infrared face recognition, the four existing features extraction algorithms are used for comparisons that include PCA+LDA [13], DCT+LDA [7], DCT+ FS+LDA [8]. We have used our own implementations of all of these algorithms on our infrared face database. The best performances of different algorithms are shown in Table 1.

Table 1. Best recognition rates of different algorithms

Methods	Results
DCT+PLS (Proposed)	95.8%
PCA+PLS	92.4%
DCT+FS +LDA[8]	93.6%
DCT+LDA[7]	91.2%
PCA+LDA[13]	89.2%

It is revealed from tabel that the recognition performance of the algorithm based on DCT and PLS is very high and outperforms that of the methods based on DCT and LDA. This is because the PLS is a powerful feature extraction technique for discriminative information in DCT domain. We observed that the PLS regression performed better than LDA because PLS basis projects the feature vectors into a latent space in which feature vectors corresponding to the same subject are closer than the feature vectors corresponding to different subjects.

5 Conclusions

In this research paper we presented a DCT based feature extraction method for the representation of infrared face images. To perform face recognition, the proposed features were classified using the PLS regression. The experiments were performed on our infrared face datasets and the results of the proposed algorithm were compared with other state-of-the art infrared face recognition algorithms based on DCT and PCA. The experimental results proved that the proposed algorithm consistently outperforms the existing methods.

Acknowledgements. While working on this research paper, we were supported by the National Nature Science Foundation of China (No. 61201456), the Natural Science Foundation of Jiangxi Province of China (No. 20132BAB201052), the Science & Technology Project of Education Bureau of Jiangxi Province (No.GJJ14581) and the Nature Science Project of Jiangxi Science and Technology University (2013QNBjRC005, 2013ZDPYJD04); we would like to show our highest respect here to all those who provided help.

References

1. Han, H., Shan, S., Chen, X., Gao, W.: A Comparative Study on Illumination Preprocessing in Face Recognition. *Pattern Recognition* 46(6), 1691–1699 (2013)
2. Jin, H., Wang, R.: Robust Image Set Classification Using Partial Least Squares. In: Sun, C., Fang, F., Zhou, Z.-H., Yang, W., Liu, Z.-Y. (eds.) *ISCIIDE 2013*. LNCS, vol. 8261, pp. 200–207. Springer, Heidelberg (2013)
3. Li, S.Z., Jain, A.K.: *Handbook of Face Recognition*, 2nd edn. Springer (2011) ISBN 978-0-85729-931-4
4. Hermosilla, G.: A Comparative Study of Thermal Face Recognition Methods in Unconstrained Environments. *Pattern Recognition* 45(7), 2445–2459 (2012)
5. Hafed, Z.M., Levine, M.D.: Face Recognition Using the Discrete Cosine Transform. *International Journal of Computer Vision* 43(3), 167–188 (2001)
6. Hongtao, Y., Ping, F., Xuejun, S.: Face Recognition Based on DCT and LDA. *Acta Electronic Sinica* 37(10), 2211–2214 (2009)
7. Zhang, Y.-K., Liu, C.-Q.: A Novel Face Recognition Method Based on Linear Discriminant Analysis. *Journal of Infrared and Millimeter Waves* 22(5), 327–330 (2003)
8. Xie, Z., Liu, G., Wu, S., et al.: A Novel Infrared Face Recognition Method in DCT Domain. In: 2010 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), pp. 12–16 (2010)
9. Choisy, J., Schwartz, W.R., Guo, H., et al.: A Complementary Local Feature Descriptor for Face Identification. In: 2012 IEEE Workshop on Applications of Computer Vision, WACV, pp. 121–128 (2012)
10. Schwartz, W.R., Guo, H., Davis, L.S.: A Robust and Scalable Approach to Face Identification. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010, Part VI*. LNCS, vol. 6316, pp. 476–489. Springer, Heidelberg (2010)
11. Sharma, A., Jacobs, D.: Bypassing Synthesis: PLS for Face Recognition with Pose, Low-Resolution and Sketch. In: 2011 International Conference on Computer Vision and Pattern Recognition, CVPR, pp. 593–600 (2011)
12. Partial Least Square Tutorial, <http://www.statsoft.com/textbook/partial-least-squares/#SIMPLS>
13. Hua, S.G., Zhou, Y., Liu, T.: PCA+LDA Based Thermal Infrared Imaging Face Recognition. *Pattern Recognition and Artificial Intelligence* 21(2), 160–164 (2008)
14. Uzair, M., Mahmood, A., Mian, A.: Hyperspectral Face Recognition using 3D-DCT and Partial Least Squares. In: 2013 British Machine Vision Conference, BMVC 2013, pp. 1–9 (2013)

Pipeline Architecture for High Speed License Plate Character Recognition

Boyu Gu¹, Qiang Zhang², and Zhenhuan Zhao²

¹ Changchun University of Science and Technology
No.7089, Weixing Road, Changchun, 130022, China
guboyu1101@163.com

² Continental Automotive Corporation (Lian Yun Gang) Co. Ltd. Changchun Branch
No.1981, Wuhan Road, Changchun, 130000, China

Abstract. An embedded hardware for license plate character recognition is designed and implemented on an FPGA (field programmable gate array) with pipeline architecture. The architecture is based on M2DPCA (modular two-dimensional principal component analysis) algorithm. Three processing elements are contained in the proposed pipeline architecture, projection element is designed for matrix multiplication operations of feature extraction, the distances between input character and each class in training database are computed in distance element, and the nearest neighbor classification is carried out in classification element, all functions are run in pipeline. Experimental results show that very high speed is achieved, which provides approximately 28% speedup of equivalent software implementation, and also, the hardware architecture performs extremely resource economical.

Keywords: License plate recognition, character recognition, FPGA, pipeline processing, hardware architecture.

1 Introduction

Automatic license plate recognition technology has numerous important applications in people's daily life[1,2]. In a license plate recognition system, very little time and resource is allowed to be consumed by character recognition. One main difficulty of license plate character recognition is that implementations on embedded application should operate fast enough to ensure the whole system to fulfill the real-time requirement[3], and the other obstacle is make the resource occupation of character recognition functions as little as possible.

To efficaciously extract the feature in a high dimensional space is extremely crucial for character recognition. Since high dimensional image data could projected into low dimensional eigen space, the PCA based approach is very suitable for embedded applications. As a statistical approach, PCA was proposed in [4], and has been widely applied to pattern recognition[5,6]. To improve the accuracy to varying illumination and angle, modular PCA was proposed in [7], by divide the images into sub-blocks, the technique has been achieve a better recognition rate. 2DPCA was proposed in [8],

which is based on 2D image matrixes rather than 1D vectors, and the computational complexity is significantly reduced. And M2DPCA was proposed in [9], by combining the strengths of both modular PCA and 2DPCA, which performances more efficient and robust.

In recent years, many researchers have taken the advantages of FPGAs to implement character recognition for practical applications due to its strengths such as low power consumption, capability to create customizable portable devices, and foremost, high performance can be achieved by means of applying the embedded memory modules and DSP units. Depending on the pipeline and parallel processing, real-time license plate character recognition can be achieved.

This paper focus on achieve an embedded real-time license plate character recognition architecture. Three kinds of processing element are designed, the projection element and distance element are explored to operate the data of image sub-blocks, and the classification element is presented for nearest neighbor classification. The arithmetic and logic functions are described in Verilog HDL. Since the processing elements can be operated in pipeline, FPGA implementations for character recognition based on M2DPCA could achieve a remarkable high speed.

The rest of this paper is organized as follows. In Section 2, the M2DPCA algorithm for character recognition is discussed. In Section 3, the design of hardware architecture is illustrated in detail. In Section 4, experimental results are presented. In Section 5, Conclusions and ideas for future work are given.

2 Review of M2DPCA

In M2DPCA, each sub-block of character image is represented by a matrix in eigen space, and new character image is classified corresponding to the closest matching sample in training database.

2.1 Training Phase

Suppose that p classes are included in training database, each class contains q images, every image is divided into s sub-blocks ($s=b^2$), and the size of images is $m \times n$. Every image is a training sample, each sub-block is regarded as an $m/b \times n/b$ image matrix, and $p \times q \times s$ matrix are included in the sample space I . The j th sample of i th class in I is expressed as:

$$\mathbf{I}_{ij} = \begin{Bmatrix} \mathbf{I}_{ij,11} & \mathbf{I}_{ij,12} & \cdots & \mathbf{I}_{ij,1b} \\ \mathbf{I}_{ij,21} & \mathbf{I}_{ij,22} & \cdots & \mathbf{I}_{ij,2b} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{I}_{ij,b1} & \mathbf{I}_{ij,b2} & \cdots & \mathbf{I}_{ij,bb} \end{Bmatrix} \quad (1)$$

The average of all image matrixes in training sample space is computed as:

$$\boldsymbol{\mu} = \frac{1}{pqs} \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^s \mathbf{I}_{ijk} \quad (2)$$

Use the average to centralize every image matrix, and compute the covariance matrix of sample space:

$$\mathbf{C} = \frac{1}{pqs} \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^s (\mathbf{I}_{ijk} - \boldsymbol{\mu})(\mathbf{I}_{ijk} - \boldsymbol{\mu})^T \quad (3)$$

Compute all the orthonormalized eigenvectors of covariance matrix \mathbf{C} and corresponding eigenvalues, then, the eigenvectors corresponding to largest ε eigenvalues are elected to form a matrix \mathbf{E} , the size of \mathbf{E} is $m/b \times \varepsilon$, and $\mathbf{E} \times \mathbf{E}^T$ is an identity matrix. The projection of a matrix form sample space to eigen space is:

$$\boldsymbol{\delta}_{ijk} = \mathbf{E}^T (\mathbf{I}_{ijk} - \boldsymbol{\mu}) \quad (4)$$

Where \mathbf{I}_{ijk} is the (ijk) th image matrix in sample space, that is also the k th sub-block of j th sample in i th class, and $\boldsymbol{\delta}_{ijk}$ is the corresponding projection matrix in eigen space. The mean projection of a sub-block in same class is computed as:

$$\boldsymbol{\eta}_{ik} = \frac{1}{q} \sum_{j=1}^q \boldsymbol{\delta}_{ijk} \quad (5)$$

Every $m/b \times n/b$ dimensional matrix in sample space represent as a $\varepsilon \times n/b$ dimensional matrix in eigen space, $\varepsilon \leq m/b$ (generally ε is much less than m/b).

2.2 Classification Phase

A character image needs to be recognized is a considered as a test sample, the k th sub-block of test sample is a matrix $\mathbf{I}_{test,k}$, the low dimensional projection of which in eigen space is:

$$\boldsymbol{\eta}_{test,k} = \mathbf{E}^T (\mathbf{I}_{test,k} - \boldsymbol{\mu}) \quad (6)$$

Euclidean distance from \mathbf{I}_{test} to i th class in training database is computed as:

$$d_i = \frac{1}{s} \sum_{k=1}^s \|\boldsymbol{\eta}_{test,k} - \boldsymbol{\eta}_{ik}\| \quad (7)$$

At last, the test sample could be classified to Γ th class when:

$$\Gamma = \operatorname{argmin}(\mathbf{D}) \quad (8)$$

Where \mathbf{D} is a set composed by the distances between test sample and all classes, $\mathbf{D} = \{d_1, d_2, \dots, d_p\}$.

3 Hardware Architecture Design

3.1 Modified Distance Equation

A modified distance equation is introduced to simplify the operations of M2DPCA eigen space projection for hardware implementation. Based on Equation (4)~(6), the distance between input character and i th class on k th sub-block can be computed as:

$$\begin{aligned}
 d_{ik} &= \left\| \mathbf{E}^T (\mathbf{I}_{test,k} - \boldsymbol{\mu}) - \frac{1}{q} \sum_{j=1}^q \mathbf{E}^T (\mathbf{I}_{ijk} - \boldsymbol{\mu}) \right\| \\
 &= \left\| \mathbf{E}^T \mathbf{I}_{test,k} - \frac{1}{q} \sum_{j=1}^q \mathbf{E}^T \mathbf{I}_{ijk} \right\| = \left\| \mathbf{E}^T \mathbf{I}_{test,k} - \gamma_{ik} \right\|
 \end{aligned}
 \tag{9}$$

In Equation (9), $\boldsymbol{\mu}$ is removed, and γ_{ik} is a constant which can be obtained in training phase because of all needed variables are available off-line. In the proposed architecture, all operations in training phase are run on a PC. And in Equation (7), since s is a constant, it can be ignored in hardware Implementation.

3.2 Structure of Processing Elements

The core of proposed hardware architecture is the processing elements. The projection element is designed to project a sub-block of input character into eigen space, which is essentially a matrix multiplier. The function of distance element is computing the Euclidean distances between input character and each class in training database. And nearest neighbor classification is carried out in the classification element. The structure of processing elements is shown in Fig. 1. The k th sub-block is described in detail, and others with same structure are omitted.

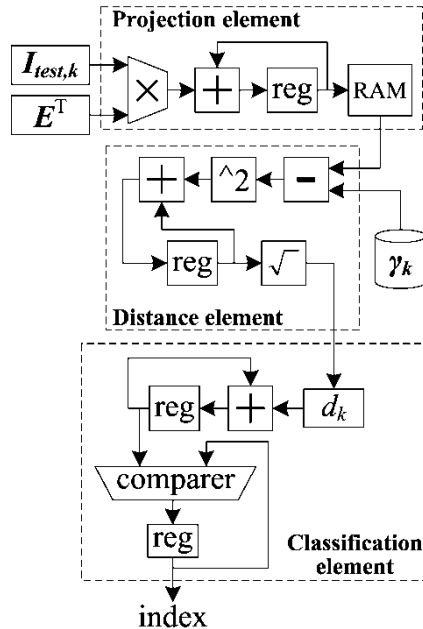


Fig. 1. Structure of processing elements

Matrix multiplication is actually a series of multiplication and addition, which can be performed by a multiplier, an adder, and a register in the projection element. The result is stored in an on-chip RAM (random access memory) instead of logic cell based registers. Since amount of memory bits in FPGA is much greater than logic cells, the matrix operation is more resource economical.

Euclidean distance is computed in the distance element. The square operated by a multiplier, the two inputs of which are connected to same data source, or a LUT (lookup table) is also practicable. The square root is computed by a LUT, but in fact, the square root is not necessary in some circumstances.

In the classification element, an adder and a register are used to summate distances of all sub-blocks. The minimum distance and corresponding class index is computed by the comparer.

3.3 Frame of Pipeline Architecture

In consideration of both of recognition speed and resource consumption, the processing elements are organized in the pipeline architecture. Data obtained in training phase is stored in ROMs (read only memory), which includes the eigenvectors and low dimensional eigen space projections of training database. The frame of pipeline architecture is shown in Fig. 2.

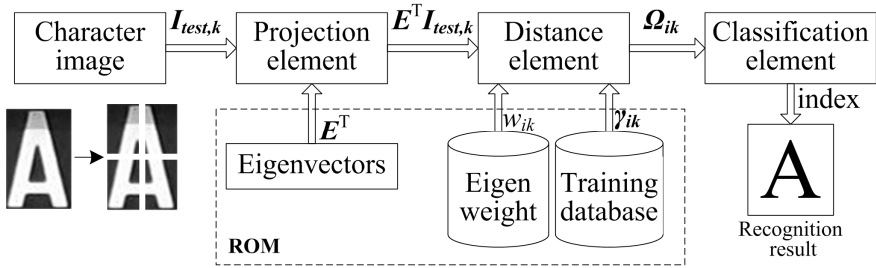


Fig. 2. Frame of pipeline architecture

3.4 Timing Analyzes

The timing sequence of pipeline architecture is shown in Fig. 3. When an input character is divided into s sub-blocks, there are $s+1$ periods contained in the recognition procedure. In the first period, only the projection element is active, and in last period, only the projection element is idle, and the comparer of classification element for minimum distance determination is enabled. Except the first and last period, all other periods are identical, and the k th period is demonstrated in detail.

Since the size of E^T is $\epsilon \times m/b$, for each sub-block, $m \times n \times \epsilon/s$ cycles are needed to project the input character into eigen space, and the arithmetic and logic components lead to 3 cycles latency. Because of all projections in eigen space is a $\epsilon \times n/b$ dimensional matrix, the distance element takes $\epsilon \times n \times p/b$ cycles to compute distances of same sub-block between input character and p classes in training database, and 6 cycles latency is involved. The classification element summates distances of all

sub-blocks together in each of the 2th~(s+1)th period. In last period, the minimum distance is obtained in p cycles and cause 2 cycles latency. Since the processing elements are run in pipeline, the serial architecture takes $m \times n \times \varepsilon + \varepsilon \times n / b \times p + p + 12$ clock cycles to recognize a character. There is idle time in projection element if $p > m/b$, and vice versa for distance element.

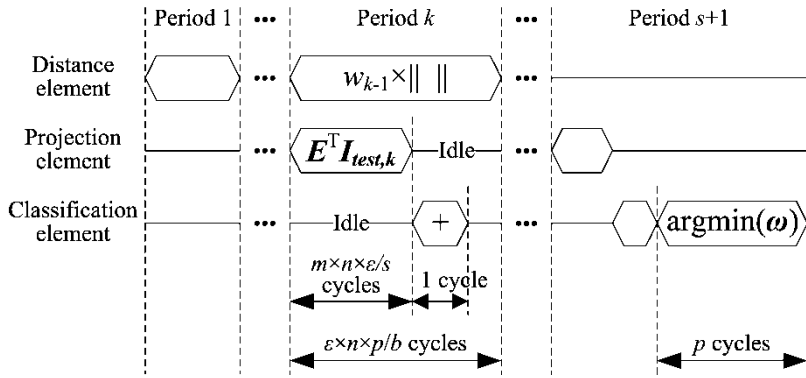


Fig. 3. Timing sequence of pipeline architecture

4 Experimental Results

The performance of the hardware architecture is tested on a character database based on Chinese license plate, Fig. 4 shows some examples in the character database.



Fig. 4. Examples in character database

The database includes 2600 grayscale images in total. All the images are scaled to 24×48 . The database consists of 65 classes, include 31 Chinese characters, 24 upper case English letters (the "I" and "O" are not contained in any Chinese license plate),

and 10 Arabic numbers, each class contains 40 samples. In praxis, characters images are divided into 2×2 sub-blocks, the size of each sub-block is 12×24 , 20 images of each class in the database are used for compose the training set, and others are used for testing.

The hardware architecture is implemented with Verilog HDL utilizing Quartus II synthesis software, the target FPGA is an Altera Cyclone II chip which contains 68,416 logic cells, 1,152,000 memory bits, and 150 embedded multipliers, and the width of eigen space projection is 32 bit. Recognition rate of varying number of eigenvectors (1~12) is shown in Table 1. The optimum recognition rate is 96.77%, and the corresponding number of eigenvectors is 6.

Table 1. Recognition rate

Eigenvectors	Recognition rate (%)
1	66.77
2	93.54
4	96.38
6	96.77
8	96.46
10	95.92
12	96.15

In practical implementation, ε is set to 6 for feature extraction. Table 2 shows resource consumption of the hardware architecture, occupies about 36% on-chip memory bits, and 11% embedded multipliers. With all other functions (image input/output, memory and synchronization control, etc.), only 6% logic cells are consumed in total.

Table 2. Resource consumption

	Consumed	Total on-chip
Logic cells	3782	68,416
Memory bits	412,176	1,152,000
Embedded multipliers	34	300

All character recognition functions in the pipeline architecture are run at 100 MHz, compare it against the software which is implemented on an AMD dual-core 2.6 GHz PC with Matlab 7.6. The recognition speed is shown in Table 3.

Table 3. Recognition speed

	Hardware	Software
Target device	EP2C70F896C6	PC
Synthesis tool	Quartus II 12.0	Matlab 7.6
Total clock cycles	21,322	--
Clock frequency	100 MHz	--
Recognition time	213.2 μ s	297.4 μ s

In the pipeline architecture, the projection element takes 6988 cycles, the distance element needs 19,504 cycles, and the minimum obtained in 68 cycles. The total recognition time is much less than the summation of time consumed by each processing elements because of all the functions are pipelined. The recognition time is not fully match the timing analyze (Section 3.4) since the time delay such as memory addressing and image inputting are involved in practical application. The pipeline architecture is capable of recognizing 4690 characters in one second. A Chinese license plate contains 7 characters, and it takes about 1.5 ms to recognize a license plate. As shown in Table 3, the recognition speed of software implementation is 297.4 μ s, and the pipeline architecture is much faster, which provides about 28% speedup.

5 Conclusions

High performance real-time hardware architecture for license plate character recognition was designed and implemented on an FPGA. Pipeline architecture was explored, and significant speedup over equivalent software implementation achieved. Experimental results indicate that FPGAs are very suitable for PCA based character recognition.

Although the hardware architecture was tested on a database of Chinese license plates, it can be modified to adapt other license plates. The recognition speed of the hardware is related to the clock frequency, even we did not have a device for testing, there is no reason that a fast hardware would not be achieved on FPGAs with faster speed grade. In future work, entire license plate recognition system will be implemented on an FPGA.

References

1. Huang, Y.S., Weng, Y.S., Zhou, M.C.: Critical scenarios and their identification in parallel railroad level crossing traffic control systems. *IEEE Transactions on Intelligent Transportation Systems* 11(4), 968–977 (2010)
2. Omitaomu, O.A., Ganguly, A.R., Patton, B.W., Protopopescu, V.A.: Anomaly detection in radiation sensor data with application to transportation security. *IEEE Transactions on Intelligent Transportation Systems* 10(2), 324–334 (2009)
3. Anagnostopoulos, C.N.E., Anagnostopoulos, I.E., Psoroulas, I.D., Loumos, V., Kayafas, E.: License plate recognition from still images and video sequences: A survey. *IEEE Transactions on Intelligent Transportation Systems* 9(3), 377–391 (2008)
4. Kirby, M., Sirovich, L.: Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Analysis and Machine Intelligence* 12(1), 103–108 (1990)
5. Turk, M., Pentland, A.: Face recognition using eigenfaces. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–591 (1991)
6. Zhang, D., Mabu, S., Hirasawa, K.: Robust intelligent PCA-based face recognition framework using GNP-fuzzy data mining. *IEEJ Transactions on Electrical and Electronic Engineering* 8(3), 253–262 (2013)

7. Gottumukkal, R., Asari, V.K.: An improved face recognition technique based on modular PCA approach. *Pattern Recognition Letters* 25(4), 429–436 (2004)
8. Yang, J., Zhang, D., Frangi, A.F.: Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence* 26(1), 131–137 (2004)
9. Chen, F., Chen, X., Zhang, S., Yang, J.: A Human Face Recognition Method Based on Modular 2DPCA. *Image and Graphics* 11(4), 580–585 (2006)

Robust Dual-Kernel Tracking Using Both Foreground and Background

Wangsheng Yu, Zhiqiang Hou, Xiaohua Tian, Lang Zhang, and Wanjun Xu

Information and Navigation College, Air Force Engineering University, Xi'an, China
xing_fu_yu@sina.com

Abstract. The kernel-based mean shift tracker outperforms other trackers due to its innovated target representation and efficient optimization strategy. However, this representation relies overmuch on the foreground and thus, decreases the robustness to the background change and clutter. To this point, this paper presents a dual-kernel tracker based on mean shift using both foreground and background. The proposed target representation consists of foreground model and background model, and the optimizing process integrates foreground kernel iteration and background kernel iteration. Experiments indicate that the proposed tracker obtains better performance in coping with background change and clutter.

Keywords: Visual tracking, mean shift, kernel-based tracker, dual-kernel tracker.

1 Introduction

Visual tracking is widely used in civil and military fields. As one of the famous trackers, the kernel-based tracker [1] is essentially a model-driven tracker based on an efficient searching algorithm. It searches the local maxima along with the ascent direction of gradient in feature space [2], which is known as mean shift iteration. Due to its simplicity and efficiency, mean shift algorithm has been widely applied in visual tracking, image smooth, cluttering and segmentation [3]. Another technique presented in kernel-based tracker is target representation. It spatially masks the target window with an isotropic kernel and transforms it into histogram features. In the past decade, many researchers did further studies to improve the tracking performance of kernel-based tracker. The fruitful works varies from kernel bandwidth selection [4], spatial histogram [5], adaptive binning histogram [6], anisotropic kernel [7], background contrasting [8], multi-part model [9], to more discriminative similarity metric [10]. The common point shared among these works is that a single kernel and limited background information is utilized.

In this paper, we proposed a dual-kernel tracker using both foreground and background, that is, both foreground kernel and background kernel are utilized to accomplish the tracking task. The fully utilizing of sufficient background information obviously improved the tracking accuracy and robustness.

2 Traditional Kernel-Based Tracker

The kernel-based tracker describes the target model with m -bins histogram features. Let $\{x_i\}_{i=1\dots n}$ be the pixel location centered at x_0 and $\mathbf{q} = \{q_u\}_{u=1\dots m}$ be the normalized target model. The sub-feature q_u in this model can be computed as:

$$q_u = c_q \sum_{i=1}^n K_h(x_i - x_0) \delta[b(x_i) - u] \quad (1)$$

where c_q is a normalizing constant and $K_h(x)$ is the kernel function with bandwidth of h , which will be discussed later. δ is the Kronecker delta function and $b(x_i)$ is the feature mapping function from the color of location x_i to the histogram bins.

It calculates the sub-features of target candidate $\mathbf{p}(y) = \{p_u(y)\}_{u=1\dots m}$ centered at y in the same way,

$$p_u(y) = c_{py} \sum_{i=1}^n K_h(x_i - y) \delta[b(x_i) - u] \quad (2)$$

where c_{py} is a normalizing constant. Then Bhattacharyya coefficient is utilized to calculate the similarity between target model and candidate,

$$\rho(y) \equiv \rho[\mathbf{p}(y), \mathbf{q}] = \sum_{u=1}^m \sqrt{p_u(y) \cdot q_u} \quad (3)$$

Expanded around y_0 using the first-order Taylor series and integrated with formula (2), the former equation can be approximated as

$$\rho(y) \approx \frac{1}{2} \sum_{u=1}^m \sqrt{p_u(y_0) \cdot q_u} + \frac{c_{py}}{2} \sum_{u=1}^m \omega(x_i) K_h(y - x_i) \quad (4)$$

where

$$\omega(x_i) = \sum_{u=1}^m \sqrt{q_u / p_u(y_0)} \delta[b(x_i) - u] \quad (5)$$

Then the mean shift algorithm is utilized to optimize the equation (4) and the kernel is then recursively moves from the current location y_0 to the new location y_1 according to the following iteration:

$$y_1 = \frac{\sum_{i=1}^{n_h} G_h(x_i - y_0) \omega(x_i) x_i}{\sum_{i=1}^{n_h} G_h(x_i - y_0) \omega(x_i)} \quad (6)$$

where $G_h(x) = -K'_h(x)$. The procedure ends when the iteration time is up to a preset threshold or the distance between two consecutive results (stopping criterion) is under another preset threshold.

The tracker chooses Epanechnikov kernel as kernel function, which assigns a bigger weight to the locations nearer by the center of the target. The assigned weights make the probability density function smoother and enhance the robustness of the tracker. However, it excludes the background information which is important to tracking task. Some subsequent works are centralized on this problem, but almost all the ameliorations are to correct the model features by suppressing the background information.

3 The Proposed Dual-Kernel Tracker

In this paper, we take the advantage of background information in a different way. Two kernels are designed to respectively focus on the foreground and background. The tracker based on these two kernels obtains a higher precision and reveals the robustness to the background change and clutter.

Unlike the kernel-based tracker, we choose Gaussian kernel as the kernel functions and describe them as follows:

$$K_h^{fg}(x) = c_{fg} \exp\left(-\frac{1}{2} \left\| \frac{x}{\lambda \cdot h} \right\|^2\right) \quad (7)$$

$$K_h^{bg}(x) = c_{bg} \left[\max(K_h^{fg}) - K_h^{fg}(x) \right] \quad (8)$$

where $K_h^{fg}(x)$ is foreground kernel and $K_h^{bg}(x)$ is background kernel. c_{fg} and c_{bg} are the normalization constants. $\max(K_h^{fg})$ is the maximum value of $K_h^{fg}(x)$. The bandwidth h here is the window size of an enlarged region which contains both foreground and background. λ is the ratio of the target size to the enlarged size. These two parameters are utilized as follows:

$$\left\| \frac{x}{\lambda \cdot h} \right\| = \frac{1}{\lambda} \sqrt{\left(\frac{x}{h_x}\right)^2 + \left(\frac{y}{h_y}\right)^2} \quad (9)$$

where $x = (x, y)$ and $h = (h_x, h_y)$.

With these two kernels, we obtain a new iteration formula as follows:

$$y_1 = \theta_1 \cdot y_1^{fg} + \theta_2 \cdot y_1^{bg} \quad (10)$$

where y_1^{fg} and y_1^{bg} are obtained from formula (6) respectively with foreground kernel and background kernel. θ_1 and θ_2 are the weights calculated by

$$\theta_1 = c_1 \cdot \rho \left[\mathbf{p}^{fg} \left(y_1^{fg} \right), \mathbf{q}^{fg} \right] \quad (11)$$

$$\theta_2 = c_2 \cdot \rho \left[\mathbf{p}^{bg} \left(y_1^{bg} \right), \mathbf{q}^{bg} \right] \quad (12)$$

where \mathbf{q}^{fg} is the foreground model and \mathbf{q}^{bg} is the background model. $\mathbf{p}^{fg} \left(y_1^{fg} \right)$ is the foreground candidate at y_1^{fg} and $\mathbf{p}^{bg} \left(y_1^{bg} \right)$ is the background candidate at y_1^{bg} . c_1 and c_2 are utilized to normalize θ_1 and θ_2 .

A smaller θ_1 denotes the appearance variation of foreground which may due to the illumination or background change. For this condition, the tracking result from background kernel plays an important role. On the contrary, a smaller θ_2 denotes a distinct change of background and meanwhile the foreground is more reliable.

The background kernel also plays an important role in correcting model drift. We utilize the following formula to updating the background model

$$\mathbf{q}^{bg} = c \cdot \left[\frac{\mathbf{p}^{bg} \left(y_1^{bg} \right) + \mathbf{q}^{bg}}{2} \right] \quad (13)$$

where c is a normalization constant. We update the background model each iteration to adaptive the background change while relatively restrict the foreground model renew. The reason is that the background may change most of the time but the foreground is relatively constant. It should be noted that a suitable updating strategy of foreground model can largely suppress the model drift problem. However, it is beyond the discussing in this paper. We refer to some relative works [11, 12] to this problem.

4 Experimental Results

We evaluate the tracking performance of the proposed dual-kernel tracker (DKT), the kernel-based tracker (KBT) [1], and the KBT based on background contrasting (BC-KBT) [8]. For all the trackers, we select a three times bigger region around the target as background. The upper limit of mean shift iterations is set at 15 and the stopping criterion threshold is set at 0.1. Of particular note, we set the tracking window as fixed for all the trackers although the targets decreased in size.

Figure 1 shows the tracking results of an airplane taking off from the runway. In the video clip, the background changes acutely several times, which affects the tracking results of KBT and BC-KBT. Both KBT and BC-KBT lost the target in frame of 150 and 250, while the proposed DKT estimated the target position with acceptable precision.

Figure 2 shows the tracking results of a more challenging video sequence. A lot of background clutters appear in the latter part of the sequence, which greatly increases the tracking difficulty. Along with the target shrinking its size, the surrounding backgrounds changed severely, and both KBT and BC-KBT are cheated by the background clutters.

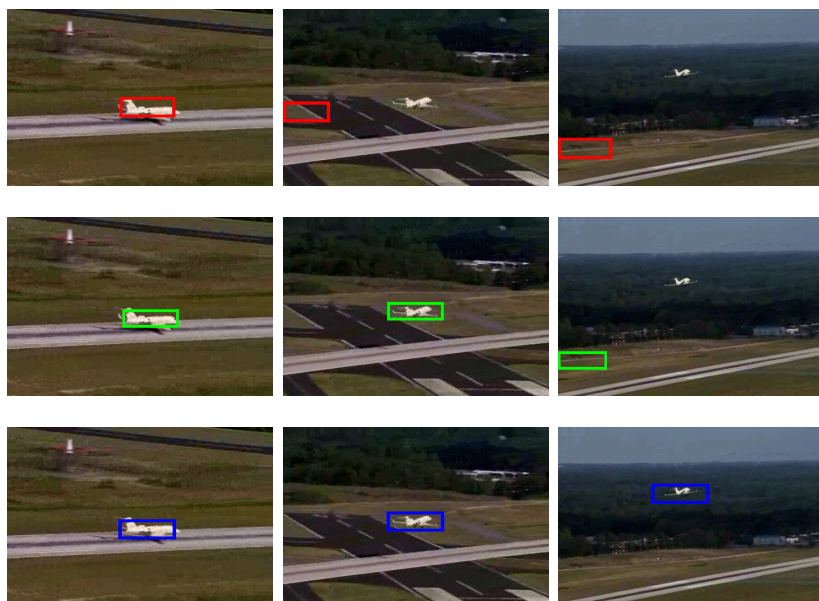


Fig. 1. Tracking evaluation on sequence with obvious background change. From top to bottom are respectively the tracking result of KBT, BC-KBT and DKT. The frame number from left to right is 50, 150, and 250.

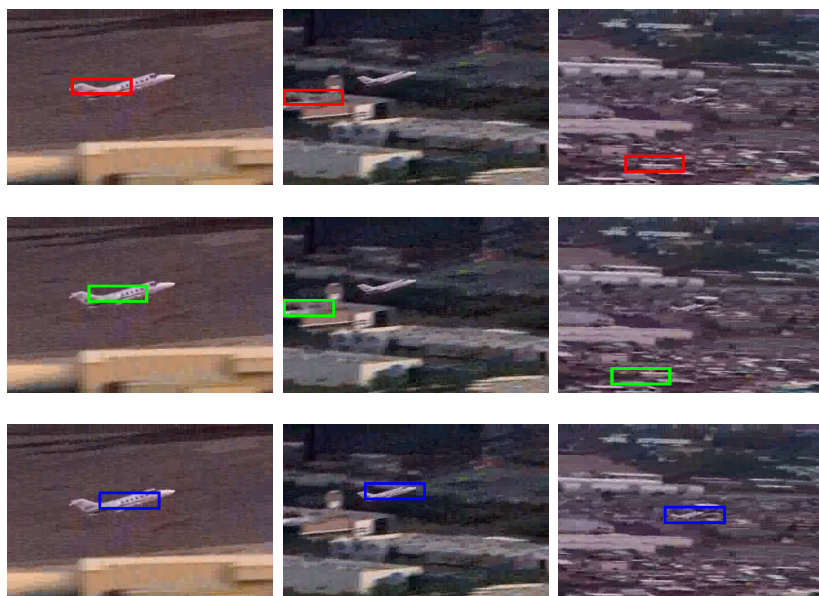


Fig. 2. Tracking evaluation on sequence with complicated background clutters. From top to bottom are respectively the tracking result of KBT, BC-KBT and DKT. The frame number from left to right is 150, 270, and 350.

The subjective sense from Figure 1 and Figure 2 indicates that the proposed tracker outperforms the traditional kernel-based tracker and its amelioration in most situations.

We quantitatively evaluate the performance using the center location error metric. It measures the Euclidean distance between the center locations between the tracking result and the ground truth. A smaller error indicates a better performance. Figure 3 shows the error plots of the trackers. Numerically, the mean errors of KBT, BC-KBT and DKT are 138.9, 137.0 and 5.4 for the first sequence, and 62.5, 59.1 and 17.9 for the second sequence. The results demonstrate that the proposed tracker exceeds the referenced trackers.

We refer to Table 1 for more details of the mean values of the center location errors comparison.

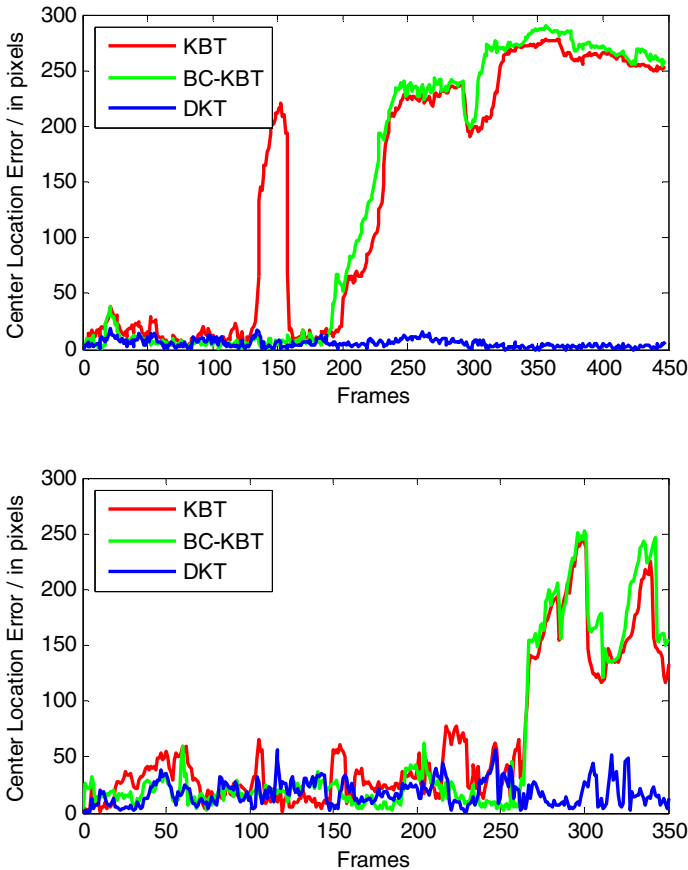


Fig. 3. The center location error plots of KBT, BC-KBT and DKT for the sequence with obvious background change (Top) and the sequence with complicated background clutters (Bottom)

Table 1. The mean value of center location errors (in Pixels)

	KBT	BC-KBT	DKT
Sequence 1	138.9	137.0	5.4
Sequence 2	62.5	59.1	17.9

5 Conclusion

In conclusion, we proposed a dual-kernel tracker based mean shift algorithm using both foreground and background. The target model consists of foreground model and background model, and the optimizing process integrates foreground kernel iteration and background kernel iteration. We update the background model each iteration to adaptive the background change, and relatively retain the foreground model to weaken the model drift. The proposed tracker is more robust in coping with the background change and clutters. The subjective tracking results and quantitative evaluation demonstrate that the proposed dual-kernel tracker outperforms the single-kernel trackers. An effective method to estimate the scale change may greatly improve the tracking performance and thus, our next work will focus the scale estimation during the iteration of two kernels.

Acknowledgements. This research was supported by National Natural Science Foundation of China (No. 61175029).

References

1. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-Based Object Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(5), 564–577 (2003)
2. Cheng, Y.: Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17(8), 790–799 (1995)
3. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(5), 603–619 (2002)
4. Collins, R.T.: Mean-shift blob tracking through scale space. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 234–240. IEEE Press, Wisconsin (2003)
5. Birchfield, S.T., Rangarajan, S.: Spatiograms versus histograms for region -based tracking. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1158–1163. IEEE Press, Santiago (2005)
6. Li, P.H.: An Adaptive Binning Color Model for Mean Shift Tracking. *IEEE Transactions on Circuits and Systems for Video Technology* 18(9), 1293–1299 (2008)
7. Qi, S., Huang, X., Yi, H.: Object tracking by anisotropic kernel mean shift. *Journal of Electronics and Information Technology* 29(3), 686–689 (2007)
8. Liu, R., Jing, Z.: Robust kernel-based tracking algorithm with background contrasting. *Chinese Optical Letters* 10(2), 021001 (2012)

9. Darren, C., Kenneth, D.H.: Evaluation of multi-part models for mean-shift tracking. In: International Machine Vision and Image Processing Conference, pp. 77–82. IEEE Press, Portrush (2008)
10. Leichter, I.: Mean shift trackers with cross-bin metrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(4), 695–706 (2012)
11. Schreiber, D.: Robust template tracking with drift correction. *Pattern Recognition Letters* 28(12), 1483–1491 (2007)
12. Peng, X., Bennamoun, M., Ma, Q., Lei, Y., Zhang, Q., Cheng, W.: Drift-correcting template update strategy for precision feature point tracking. *Image and Vision Computing* 28(8), 1280–1292 (2010)

Humanoid-Eye Imaging System Model with Ability of Resolving Power Computing

Ma Huimin¹ and Zhou Luyao²

¹Department of Electronics Engineering, Tsinghua University, Beijing, China
mhmpub@tsinghua.edu.cn

²Department of Electronics Engineering, Tsinghua University, Beijing, China
luyao.zhou.rg@gmail.com

Abstract. This paper proposes an innovative imaging system model for human eyes with resolving power calculation, which is feasible in practice and stands on solid Physical background. The model, humanoid-eye imaging system (HIS), is constructed synthesizing an imaging component model and a photo-sensing component model based on relevant parts of human eyes. HIS integrates core features and working mechanism of human eyes and can also be regarded as simulation of various real digital imaging systems. According to criteria derived from wave optics and the theory of receptors, point resolving power for HIS is defined and its calculations are deduced as functions of specified parameters of HIS and variables of object points observed by HIS. Experiment with a camera as the application of HIS show that HIS is applicable and its resolving power calculation is precise in reality. Our work supply a novel method for the first time to efficiently connect real observing conditions with computer simulation for fields related to 3D meshes management.

Keywords: Humanoid-eye imaging system (HIS), resolving power, wave optics, receptor.

1 Introduction

The simplification of 3D objects recognition procedure through simplifying objects' 3D mesh model and managing multi-resolution model rendering is one of the methods to decrease the computation complexity of 3D objects recognition and accelerate the speed of such procedure, which gain increasing concern at present.

The last few years have seen many researchers' innovative and practical progress in this field [1, 2, 3]. No matter what approaches are employed in data management and simplification, within most of these methods there is a critical step to determine the extent to which the simplification should be stopped. When applied in reality, a typical way of such step is to assign a threshold which indicates the termination of 3D meshes simplification [4]. The threshold is typically expected to reflect real situations of certain observing systems such as human eyes, cameras and so on, especially for works related to practical implementation or simulation.

Unfortunately, current approaches for simplification threshold determination lack supports from definite background and theories of physics. Most of these approaches

are hardly beyond rough estimation: some appoint a value as the resolution threshold [5, 6] and such process is simple but nevertheless has little to do with real conditions; some methods take into account some features of imaging instruments and suggest view-point oriented threshold determination algorithm, but they merely guess the form of the formula or fit with data to infer the threshold [7, 8]; some eschew the difficulties of analysis of real conditions to attain the threshold through experiments empirically[9].

In perspective of acquiring a model close to reality and intensify the credibility of the results, it is in great need to set up an approach of objects' 3D meshes simplification threshold determination, which is supported by background of reality of physics and is suitable for practical implementation in computer simulation. Relied on these principles, this paper propose a physics model -- humanoid-eye imaging system (HIS), which integrates main features of geometric optics imaging and photo- sensing components of human eyes. By Rayleigh's law and theory of receptors, the definition and criteria of point-distinguishability (that is, whether object points can be distinguished by HIS) are presented and the angle resolving power (ARP) for HIS is defined. Given both of the criteria and the definition, we subsequently calculate the view-point-dependent resolving power for HIS (in angle and length) as functions of parameters of HIS and object points' variables such as distance between object points and HIS, azimuth of object points relative to HIS, deflection of object points and speeds of object points. For real implementations, the relationship between resolving powers by wave optics and theory of receptors are evaluated and a judgment is deduced to determine resolving power calculation formula for a specific imaging system. An experiment with a digital camera as the application of HIS is described and the results are showed to demonstrate the approach's credibility and reliability in practice.

The achievement of this paper supplies an innovative means, which is close to real conditions of human eyes and other imaging systems in practice, to determine the resolving power scale factor in multi-resolution 3D mesh management and rendering for graphics-related work (specifically, objects 3D recognition and virtual reality 3D models rendering, etc.).

2 Composition and Structure of HIS

By now, physiology has gained considerable knowledge of human vision, especially in the anatomical structure of human eyes and relevant procedure such as optics imaging and photo-sensing [10]. Results from this area show that the real physical structure of human eyes is complex. But in views of imaging principle and photo-sensing process, human eyes have substantial amounts of characteristics in common with various practical optics imaging systems.

Given the abstracted imaging and sensing simplified models, an imaging system, Humanoid-eye Imaging System, is then constructed as Figure 1 shows. This system is the physics equivalent model of the whole optical imaging and sensing system of human eyes. Parameters and components in Figure 1 are explicated as following.

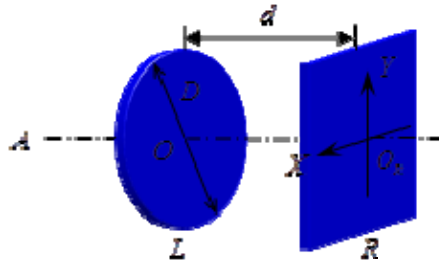


Fig. 1. System diagram of HIS

- L Imaging convergent lens (typically circular).
- R Planar sensing arrays, parallel with L , with a receptors' density function $\rho(x, y)$
- O Optical center of L .
- A Main optical axis of the system, vertical to L and R with A passing through O and O_r .
- D Diameter of L .
- d Distance between L and R .
- T Exposure time of the system.

HIS strongly focuses on the fundamental features of imaging and sensing procedure of human eyes. The working principle of HIS can be divided into imaging procedure and sampling procedure. In imaging procedure, beam from the outside world is converged by L when passing to generate a clear image on planar sensing arrays. In sampling procedure, after exposure time T , each receptor on the plane transforms the luminous energy cast on it to independently output a pixel in electronic signal and all of these pixels together compose an image output.

3 Inferring of Resolving Power for HIS

3.1 Criteria and Definition of ARP for HIS

The ability of imaging system to distinguish two object points is called the system's point resolving power (PRP). The minimum resolution angle, the minimum flair angle of two object points relative to the system when distinguishable, is often employed as the measurement of PRP and in this occasion, PRP is called angle resolving power (ARP).

According to nature of electromagnetic wave, accompanied with geometric optics are effects of wave optics, in which diffraction is a significant one. The diffraction causes image of a single object point on sensing plane to be a vague spot, Airy disk, with Airy disk's center as the ideal geometry image point. It is widely accepted that the necessary and sufficient conditions for two spots to be distinguished by an optical

lens is that the distance between two Airy disks' centers is greater than the radius of each Airy disk, which is the famous Rayleigh's law [11].

Theory of human eyes' receptors holds the viewpoint that when object points are sensible to a single receptor, there must be at least one receptor not stimulated by light between the two receptors, on which two object points' images cast separately, so as to ensure the two points are distinguishable on image. There are strong reasons behind this judgment: when image spots of two object points cast on two neighboring receptors, the system is unable to tell this situation from another one that an unique object point's image spot casts right on the boundary of the two receptors, which results in image the same as the previous situation.

Considering all described above, we define two object points distinguishable to HIS when they satisfy both following criteria.

Criterion of wave optics: The distance between centers of the two object points' image spot is not less than the larger radius of the two Airy disks.

Criterion of receptors: There is at least one un-stimulated receptor between the two receptors on which the two object points' image spots separately cast.

We define the ARP of HIS as the minimum angle of two object points relative to optical center of HIS when they are distinguishable and right satisfy one of the criterion: the distance between centers of the two object points' image spot is equal to the larger radius of the two Airy disks, or there is only one un-stimulated receptor between the two receptors where the two object points' image spots rest. By each of the criteria, two ARPs can be achieved and, naturally, we select the larger one as the actual ARP for HIS.

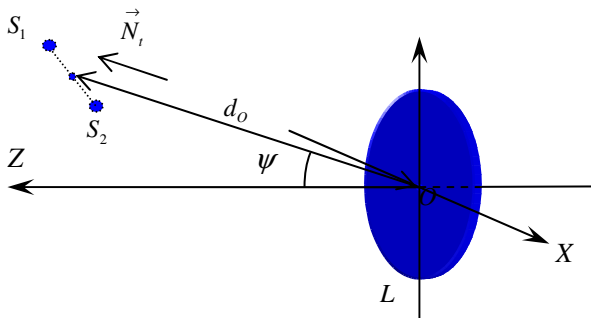


Fig. 2. Coordinator for observation of HIS

3.2 Calculation of ARP for HIS

Given object points' azimuth and speeds, parameters and principles of work of HIS (Fig.2 and Fig.3), two ARPs by both criteria are achieved. If positions and deflections of the points are available, the minimum distinguishable distance (MDD) of the points is also calculable and thus the threshold for distinguishability is established.

ARP Calculation by Criterion of Wave Optics. The criterion of wave optics requires that the distance between centers of two object points' Airy disks should be no less than the larger radius of the spots.

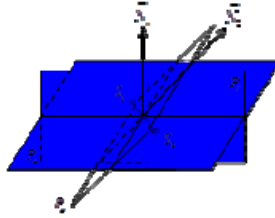


Fig. 3. Geometry relations of planes and normal vectors

By the criterion of wave optics, the ARP for HIS for static object points is the angle radius of the zero-order diffraction disk (it is easy to prove, using semi-wave-band method, that this angle radius is independent of azimuth when azimuth is not very large), that is:

$$\delta_{w01} = 1.22\lambda / D \quad (1)$$

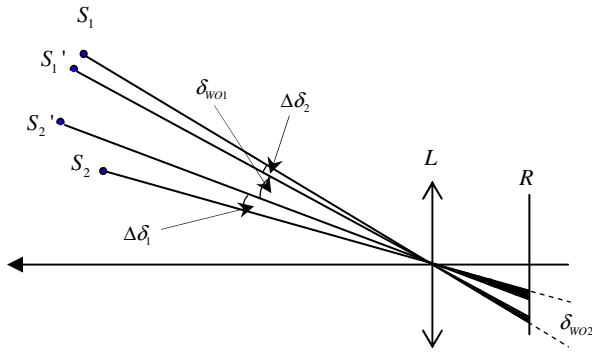
In dynamic case (the object is moving in space), however, we should consider the effects of the motion track in image on the angle of two points relative to optical center. During the exposure time, due to approaching motion, the two object points' image spots' angle relative to optical center decreases, as shown in Figure 4. The decreased angle is

$$\Delta\delta_{V_c} = \omega_c T = \frac{\|\vec{V}_c \cdot \vec{N}_\perp\| T}{d_o} \quad (2)$$

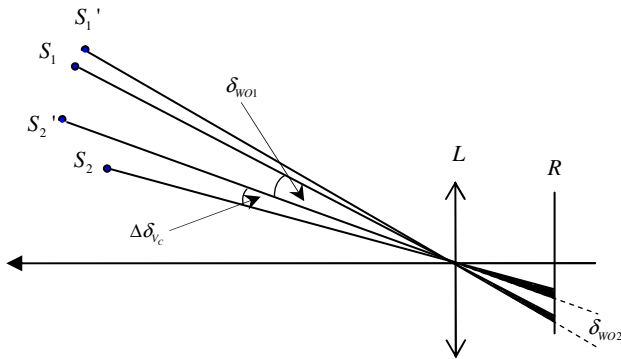
In this condition, ARP for HIS is modified to be

$$\delta_{w02} = \delta_{w01} + \Delta\delta_{V_c} = 1.22 \frac{\lambda}{D} + \frac{\|\vec{V}_c \cdot \vec{N}_\perp\| T}{d_o} \quad (3)$$

Given the positions of the two points, then the MDD along their segment is



(a) The increment of ARP is $\Delta\delta_{v_c} = \Delta\delta_1 + \Delta\delta_2 = \omega_c T = \|\vec{V}_c \cdot \vec{N}_\perp\| T / d_o$



(b) The increment for ARP is $\Delta\delta_{v_c} = \omega_c T = \|\vec{V}_c \cdot \vec{N}_\perp\| T / d_o$

Fig. 4. Effects of approaching motion on image spots' angle

$$\Delta s(d_o, \vec{V}_c) = \frac{\delta_{w02} d_o}{\|\vec{N}_s \cdot \vec{N}_t\|} = \frac{(1.22 \frac{\lambda}{D} d_o + \|\vec{V}_c \cdot \vec{N}_\perp\| T)}{\|\vec{N}_s \cdot \vec{N}_t\|} \quad (4)$$

The item $\|\vec{N}_s \cdot \vec{N}_t\|^{-1}$ is introduced as a modifier to compensate the deflection of two points relative to \vec{N}_t , as shown in Figure 5. It is obvious that the result is independent of any facts of planar sensing arrays.

ARP Calculation by Criterion of Receptors. The theory of receptors requirement for distinguishability of two object points is that there should be at least one unstimulated receptor between the two receptors on which the' image spots separately

cast. By this requirement, we evaluate effects of d_0 (distance between object points and HIS), Ψ (azimuth of object points relative to HIS), γ (deflection of object points) and V_c (speeds of object points) upon the image cast on the planar sensing arrays one by one to calculate ARP and MDD for HIS.

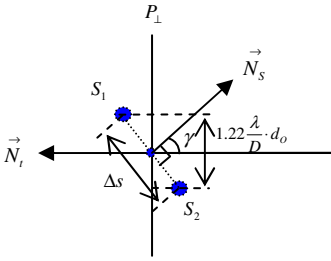


Fig. 5. Compensator $\|\vec{N}_s \cdot \vec{N}_t\|^{-1}$

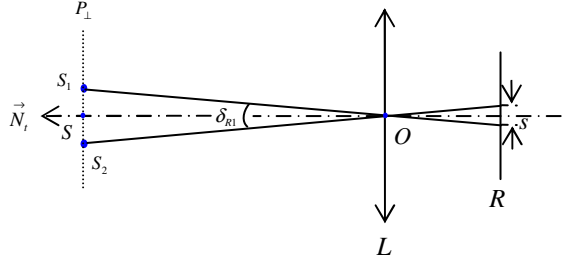


Fig. 6. Illustration of effects of d_0

i) Effect of d_0

In order to investigate on the effect of d_0 individually, we consider two static object points with $\Psi = 0$, $\gamma = 0$. In such case, d_0 and s together determine the ARP: when the angle of two point relative to optical center averagely covers right two receptor, the angle is ARP and it is easy to derive $ARP\delta_{R1}$ form the Figure 6:

$$\delta_{R1} = 2l(0,0) / d = s / d_0 \quad (5)$$

ii) Effect of Ψ

When considering effects of azimuth Ψ , facts need focus are the density variation relative to that on planar sensing arrays' origin and the length covered by the angle. Given the azimuth Ψ of two points, the length s' on the planar sensing arrays covered by the angle is (as Figure 7 illustrates)

$$s' = \theta \cdot \frac{d}{\cos \psi} \cdot \frac{1}{\cos \psi} = \frac{s}{d_0} \cdot \frac{d}{\cos^2 \psi} \quad (6)$$

When length s' is equal to double of the one-dimensional length of the local receptors, the angle is ARP:

$$s' = 2l(\psi(x_o, y_o, z_o)) = 2l\left(-\frac{x_o}{z_o}, -\frac{y_o}{z_o}\right) \quad (7)$$

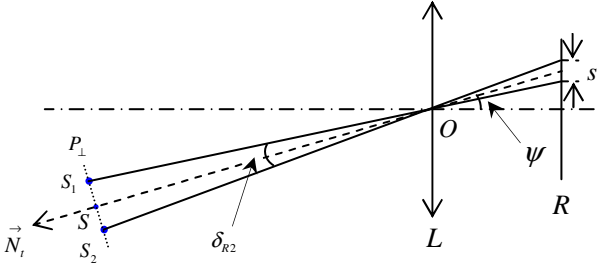


Fig. 7. Effects of azimuth ψ

Here (x_o, y_o, z_o) is the position of the geometry center of the two points. ARP δ_{R2} is then presented as

$$\delta_{R2} = 2l(\psi(x_o, y_o, z_o)) \frac{\cos^2 \psi}{d} = 2l(S_1, S_2) \frac{\cos^2 \psi}{d} \quad (8)$$

iii) Effect of \vec{V}_c

The effect of relative approaching speed \vec{V}_c is similar to that by the criterion of wave optics and the ARP δ_{R3} involving \vec{V}_c in dynamic situation is

$$\delta_{R3} = \Delta\delta_{V_c} + \delta_{P2} = \frac{\|\vec{V}_c \cdot \vec{N}_\perp\| T}{d_o} + 2l(\psi) \frac{\cos^2 \psi}{d} \quad (9)$$

3.3 Method for ARP Determination in Practice

Since the two criteria should be satisfied at the same time for HIS to distinguish two object points, both of the resolving powers by the two criteria should be calculated to find the actual resolving power for HIS. However, for most of real imaging systems, the statuses of the two criteria are not equal. Reviewing formulas for ARP and MDD calculations, we find that they are similar in structure:

$$\begin{cases} \delta_{wo} = 1.22 \frac{\lambda}{D} + \frac{\|\vec{V}_c \cdot \vec{N}_\perp\| T}{d_o} \\ \delta_R = 2l(\psi) \cdot \frac{\cos^2 \psi}{d} + \frac{\|\vec{V}_c \cdot \vec{N}_\perp\| T}{d_o} \end{cases} \quad (10)$$

$$\begin{cases} \Delta s_{wo} = (1.22 \frac{\lambda}{D} d_o + \|\vec{V}_c \cdot \vec{N}_\perp\| T) \cdot \|\vec{N}_s \cdot \vec{N}_i\|^{-1} \\ \Delta s_R = (2l(S_1, S_2) \frac{\cos^2 \psi}{d} d_o + \|\vec{V}_c \cdot \vec{N}_\perp\| T) \cdot \|\vec{N}_s \cdot \vec{N}_i\|^{-1} \end{cases} \quad (11)$$

Difference between formulas by the two criteria is the static item

$$J_{wo} = 1.22 \frac{\lambda}{D} d_o; J_R = 2l(P_1, P_2) \frac{\cos^2 \psi}{d} d_o$$

We define these two items as judgment of wave optics and judgment of receptors, which indicates resolution limit thresholds by the two criteria. Note that these judgments are determined by inherent feature parameters of HIS and by these judgments, resolving power of different applications of HIS can be assessed. For implementations in a specific imaging system, the comparison of the two judgments should be made first. Provided all outside conditions are the same, the set of formulas for ARP and MDD of which the judgment is larger is chosen as resolving power calculation for the system and the smaller one's relevant criterion will be automatically satisfied.

4 Experiments

To prove the formula of APR and MDD for HIS, an experiment is carried out with a digital camera (Canon PowerShot A710 IS) as the practical implementation of HIS and two LEDs as object points.

Relied on parameters of the camera, we calculate the judgments of the system to get $J_{wo} = 4.538 \times 10^{-5}$ and $J_R = 4.05 \times 10^{-4}$. The resolution limit threshold by theory of receptors far exceeds that by wave optics. Therefore the actual resolving power calculations for ARP and MDD are Eq. (8) and Eq. (9).

4.1 Procedure and Results

In the experiment, the effects of factors d_o , Ψ and γ are proved individually for the static situation. For dynamic situation, the effect of d_o is tested and effects of Ψ and γ are evaluated together, with different speed Vc settings for each sampling data groups. The actual approach is to calculate MDD by given parameters, then set the distance of the two LEDs to be the predicted MDD and sample images to see whether the prediction will match the results. By setting an area of 3×3 pixels as the equivalent receptor, two points are just distinguishable when the areas of 3×3 pixels where LEDs' images are located are right separated by an equivalent receptor. Some adjustment is made on original images for a better localization of LEDs. Some of the results are showed in tables 1 to 4, including original images (upper ones in each image set) and adjusted images (lower ones in each image set).

Table 1. Predicted static MDD and results for the effect of d_o




d_o / mm	1500	3300	6300
$\Delta s / \text{mm}$	3.33	7.33	14.00
image			

Table 2. Predicted static MDD and results for the effect of $\gamma(d_0 = 1800\text{mm})$




$\gamma/^\circ$	10	40	80
$\Delta s/\text{mm}$	4.06	5.22	23.04
image			

Table 3. Predicted static MDD and results for the effect of $\gamma(d_0 = 3000\text{mm})$

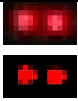




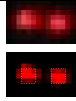
$\gamma/^\circ$	10	40	80
$\Delta s/\text{mm}$	6.77	8.70	38.39
image			

Table 4. Predicted static MDD and results for the effect of Ψ

$\Psi/^\circ$	0	+10	+20
$\Delta s/\text{mm}$	8.00	7.76	7.06
image			

4.2 Discussions

In the static part, there are 27 out of 35 (approximately 80%) images strictly satisfying the definition of just distinguishable situation defined in section 4.2. For the rest 20 percent, the error is only one pixel. Allowing for the fact that CCD samples image discretely to cause an error of a pixel for two LEDs (as discussed below), we treat these 20 percent's image falling within the error threshold and regard the prediction made by MDD formula as congruous with the whole static results. In the dynamic part, 5 out of 13 images strictly match the just distinguishable situation and for the rest images, 7 out of 8 fall in the error threshold of one pixel. We consider the dynamic results agreeable to the MDD calculation with only one image's violation.

The main source of error in the experiment comes from discretely imaging of CCD. Rather than an ideal point-photo source, the structure of LED is actually a light-emitting plane and other factors such as diffraction and diffusion also contribute to the

non-ideality. The error is at most half a pixel in situation that one edge of a LED's image falls around the center of a pixel and two LEDs' images will form an error of one pixel together at most.

Reviewing all the results, the experiment demonstrates nearly 98% images in favor of the MDD prediction and strongly supports the formula for both of ARP and MDD we achieve in chapter 3.

5 Conclusions

This paper abstracts two equivalent models of human eyes' imaging and photo-sensing parts, and then constructs HIS, an innovative imaging system which integrates main features of human eyes and many other practical imaging systems. We derive criteria of point-distinguishability for HIS and define its resolving power by Rayleigh's law and theory of receptors, which restrict the resolving power of HIS. By these criteria and definition, we propose two sets of formulas for ARP and MDD calculation, involving feature parameters of HIS and object points' variables. The experiment result shows that predictions of the resolving power calculation formulas for HIS are congruous with reality, with nearly 98% correct ratio. We have successfully simulated HIS and integrated the formulas of its resolving power in a multi-resolution object 3D models recognition computer program. The results provide supports for outstanding efficiency and applicability of HIS and relevant resolving power calculation in 3D mesh management. Future work includes evaluation of other factors related to resolving power of HIS and further application in computer.

References

1. Cignoni, P., Montani, C., Scopigno, R.: A Comparison of Mesh Simplification Algorithms. *Computers & Graphics* 22(1), 37–54 (1998)
2. Garland, M.: Multiresolution Modeling: Survey & Future Opportunities. In: *Proc. Eurographics 1999*, pp. 111–131 (1999)
3. Luebke, D.: A Developer's Survey of Polygonal Simplification Algorithms. *IEEE Computer Graphics and Applications* 21(3), 24–35 (2001)
4. Hoppe, H.: Progressive meshes. In: *Proc. SIGGRAPH 1996*, New Orleans, LA, USA, August 4-9, pp. 99–108 (1996)
5. Luebke, D.: Hierarchical structures for dynamic polygonal simplification., Technical Report, TR96.006, Department of Computer Science, University of North Carolina at Chapel Hill (1996)
6. Xia, J.C., Varshney, A.: Dynamic view-dependent simplification for polygonal models. In: *Proceedings of the IEEE Visualization 1996* (1996)
7. Hoppe, H.: View-Dependent refinement of progressive meshes. In: *Int. Proceedings of the Computer Graphics, SIGGRAPH 1997* (1997)
8. Feng, J., Zha, H.: Efficient View-Dependent LOD Control for Large 3D Unclosed Mesh Models of Environments. In: *Proc. IEEE 2004 Int. Conf. on Robotics and Automation (ICRA 2004)*, New Orleans, USA, April 26-May 1, pp. 2723–2729 (2004)

9. Murphy, H., Duchowski, A.T.: Hybrid image-/model-based gaze-contingent rendering. In: Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization (July 2007)
10. Zhang, H.: Vision and Application Technology, pp. 5–8. Zhejiang University Press, Hangzhou (2004)
11. Zhao, K., Zhong, X.: Optics, pp. 228–229. Peking University Press, Beijing (2004)
12. Shi, M.: Optics of Clinical Vision, pp. 13–14. Zhejiang University Press, Hangzhou (1993)

Color Cast Detection Method Based on Multi-feature Extraction

Minjing Miao¹, Yuan Yuan¹, Juhua Liu¹, and Hanfei Yi²

¹ School of Printing and Packaging, Wuhan University

² College of Physical Science and Technology, Huazhong Normal University
430079 Wuhan, China
christina@whu.edu.cn

Abstract. In order to raise the accuracy rate of the color cast detection and to make the method universal, the paper carries out a color cast detection method based on multi-feature extraction. Firstly, calculate the four features that are the textural property of the luminance channel, color numbers, histogram of RGB color space and statistical characteristics of the Gabor filter, then use AdaBoost to train and classify. The experiment will be done using 11346 images in the Ciurea database. The result shows that this method has a low error rate and good classification results, which is universal to natural images taken by cameras.

Keywords: Multi-feature extraction, AdaBoost; color cast detection.

1 Introduction

When capturing an image, the camera is easy to be influenced by the illumination, the reflective properties of the object itself and the photosensitive coefficient of the image capturing devices. Thus, the color of the obtained image is different from the real color of the object, which is called the image color cast [1]. And it could have a bad affection on the human visual perception. So it is important to do image color cast detection for assessing the image quality.

Nowadays the color cast detection methods are grouping into 2 parts, that are the subjective judgment and the objective classify metrics. In practice, however, subjective evaluation is usually too inconvenient, time-consuming and expensive. In the past decade, many objective color cast detection metrics have been carried out. In general, the metrics can be classified into two groups: algorithms based on the deviation of chromaticity information and algorithms that use mathematical statistics to classify the images. Examples of the first group are the White-Patch algorithm [1], the Grey-World algorithm [2], the dimensional histogram statistic algorithm [3] and the equivalent circle algorithm [4]. All above color cast detection methods are based on specific imaging assumptions. These assumptions include the set of possible light sources, the spatial and spectral characteristics of scenes, or other presumptions (e.g. white patch, averaged color is grey, etc.). As a consequence, no algorithm can be considered as universal.

Methods based on prior knowledge [5] are examples of the latter group. Such methods need to store prior knowledge to help classification. Similar approaches include methods based on machine learning [6]. But the existing methods cannot tell the color cast images apart exactly.

Therefore, in this paper, we choose AdaBoost to do the training and classification. And based on the theory of color constancy, we bring up with a color cast detection method based on multi-feature extraction.

The paper is organized as follows: In section 2, the method based on AdaBoost and feature extraction is discussed. In section 3 and section 4, the experiments are carried out on the Ciurea database. Finally, in section 5, the conclusion is provided.

2 Color Cast Detection Method Based on AdaBoost

Mankind has the ability to correct the color cast in the scene adaptively, which causes the object to be perceived constant along with the changing illuminant. And the ability is called the theory of color constancy. Taking this knowledge into consideration, we found that the color cast is not only related to the average and the variance of the chromaticity information, but also linked to the distribution of the chromaticity information [4]. In this paper, we will extract four kinds of features and then use the AdaBoost algorithm to train and classify. Fig.1 shows the workflow of the method.

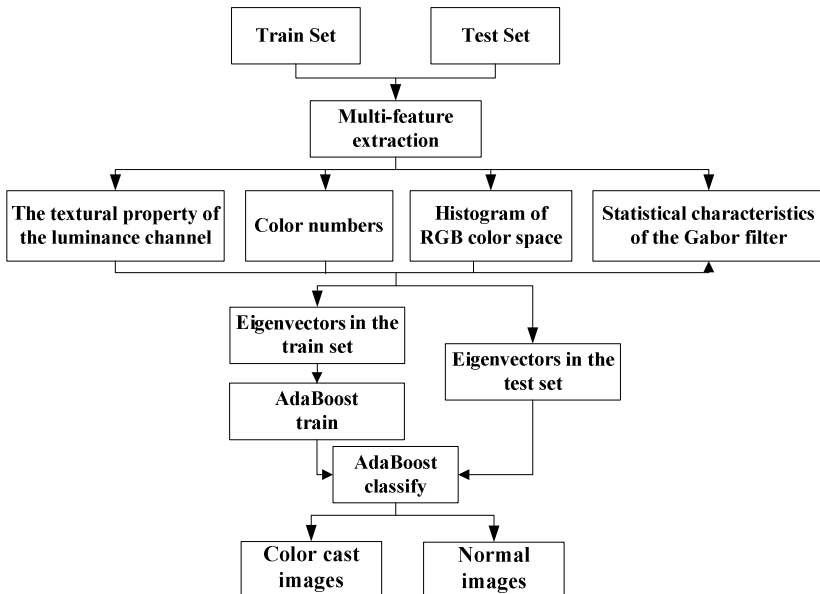


Fig. 1. Workflow of the method

2.1 Multi-feature Extraction

A. Textural Property of the Luminance Channel

When perceiving the outside world, we treat chromatic and achromatic data independently [7]. Based on this knowledge, we choose a principal component analysis (PCA [8]) to decorrelate the RGB representation of the input image into three principal components. The first component represents the largest share of signal energy. The second and the third components have one zero-crossing and two zero-crossing.

After obtaining the luminance images, we choose contrast, energy, correlation, uniformity in the gray-level co-occurrence matrix (GLCM) to represent the texture property. With a distance of 1 between the interested pixel and its neighbors, we calculate the GLCM features of 4 orientations (0° , 45° , 90° , 135°) and then sum them up to get 4 eigenvectors.

B. Color Numbers

In order to simplify the operations and to keep the color details mostly, we quantify the color levels to [0, 64), the numbers of the image color is less than $64 \times 64 \times 64$. We calculate the color numbers of the 64 scales test image and its first-derivative picture.

C. Histogram of RGB Color Space

To get the histogram feature of RGB color space, we divided each axis into four equal parts, so that the entire RGB space is broken down into total 64 parts. Each color in the image can be classified into one of the 64 parts. That is to say, we will get a 64-dimensional vector.

D. Statistical Characteristics of Gabor Filter

Gabor filter can be used to simulate the primary visual cortex receptive field properties [9]. So we will do 5 scales and 8 orientations Gabor filter in each channel of the YCbCr color space, then, compute the sum of the 40 filters and calculate the average and variance of each channel.

2.2 AdaBoost Method

AdaBoost algorithm [10,11] can be broken down into two stages. First of all, the algorithm will train a basic classifier (weak learner) for different training sets, and then put them together to make it a stronger final classifier (strong classifier).

3 Material and Method

We choose all 11346 images in the Ciurea image database [12] to do our experiment. The database which is proposed by the team of Funt in 2004 includes many kinds of natural images in our daily life and the images are with the size of 360×240 .

3.1 Prepare for the Experiment

The experiment will be distributed into 3 stages, which is shown in Fig. 2.

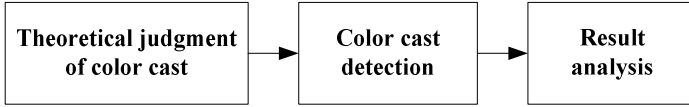


Fig. 2. Workflow of the experiment

We choose the combination of subjective judgment and the illuminant error angle to get the theoretical value of each image because of that the illuminant error angle is directly proportional to the degree of color cast. So the first stage of the experiment can be divided into 3 steps that are calculating the illuminant error angle of all images, finding out the just noticeable difference of illuminant error angle and finally choosing the value to give all image in the database a result of color cast or not.

Among the procedure 2, the just noticeable difference value can be obtained as follows:

1. We use the assumption of Von Kries in a converse way to get the RGB of the illuminant. And then calculate the illuminant error angle with Eq.1.

$$\varepsilon = \cos^{-1}[(e_i * e_e) / (\|e_i\| * \|e_e\|)]. \quad (1)$$

In this equation, e_i is the RGB of the standard illuminant and e_e is the estimated RGB of the color cast illuminant.

2. Choose about 20 standard observers in their 10° viewing angle to tell apart the chosen image is color cast or not.
3. Narrow the range of the illuminant error angle gradually and repeat the step 2 until the just noticeable difference of illuminant error angle is found.

3.2 Statistics Method

Applying the method raised in the paper, we can get an 82-dimensional eigenvector of each image. And then using the AdaBoost algorithm to train and classify the eigenvectors. After the AdaBoost classification, we can get the result of all test images. In this paper, we choose measurement ratio R_D and false alarm ratio R_F to assess the efficiency of the algorithm, and choose the misjudgment ratio to assess the iteration result. The equations are shown in (2), (3), and (4).

$$R_D = \frac{N_D}{N_+} \times 100\%. \quad (2)$$

$$R_F = \frac{N_F}{N_-} \times 100\%. \quad (3)$$

$$R_E = \frac{N_T}{N} \times 100\%. \quad (4)$$

As mentioned in (2),(3),(4), N_D is the color cast images which is also be detected as the color cast images. N_F is the normal images which has been detected as the color cast images. N_T is the sum of the error classification set and N_+ is the numbers of color cast images in the database and N_- is the numbers of normal images in the database. N is the total numbers of the database, and that is 11346.

From the equations we can see, all values are between $[0, 1]$. If R_D is higher, the classification result is better. But to R_F and R_E , the value is the smaller the better.

4 Result and Discussion

4.1 Experiment A

We choose the odd numbers of images as the train set, and the others are included in the test set. The accuracy rate of common methods can be seen in Tab.1.

Table 1. Accuracy rate using different color cast detection methods

color cast detection method	R_D	R_F	R_E
Equivalent circle algorithm	0.7821	0.9484	0.3899
FCM	0.5670	0.3591	0.1190
Proposed method	0.9753	0.1198	0.0471

From Tab.1, we can conclude that the method proposed in the paper is better than the equivalent circle algorithm and method using FCM. Because the normal image is actually rare in the database, so the false alarm ratio is a little higher than we thought. What's more, the equivalent circle algorithm is easy to be influenced by the dominant hue and the FCM method is not suitable for 2 parts classification.

4.2 Experiment B

In order to prove the effectivity of the proposed method, we have done experiment B. In the experiment, we randomly divide the database into 2 equal groups and use the two groups to train and classify. After 460 times of iterations, the result can be seen in Tab.2.

Table 2. Color cast detection result with different samples

No	R_D	R_F	R_E
1	0.9694	0.1204	0.0557
2	0.9735	0.1100	0.0462
3	0.9753	0.1198	0.0471
4	0.9725	0.1531	0.0515
5	0.9674	0.1497	0.0546

As shown in Tab.2, the proposed method is stable with good accuracy rate.

4.3 Experiment C

In this part, we have done all tests with different combination of the features raised in this paper. The good classification results with different combination of features are

as follows: 1. Histogram of RGB color space. 2. Statistical characteristics of the Gabor filter. 3. Combination of histogram of RGB color space and statistical characteristics of the Gabor filter. 4. All features combined

Tab. 3 is the result of the values of the statistics methods under 460 iterations.

Table 3. Classification results of different feature extraction methods with Adaboost

No.	R_D	R_F	R_E
1	0.9707	0.1257	0.0520
2	0.9320	0.3338	0.1306
3	0.9753	0.1198	0.0471
4	0.9735	0.1100	0.0462

From Tab. 3, we can see that all values of proposed methods are higher than 93% and the combination of all features in the paper is the best. We choose the first twome-thods in Tab. 3 to do the iteration test. The tendency chart is shown in Fig. 3.

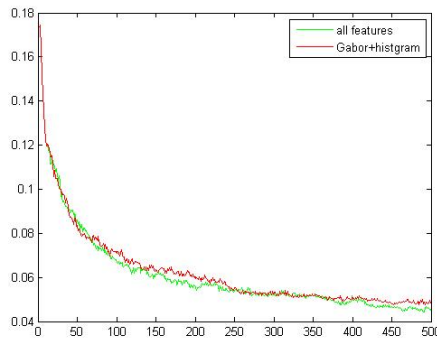


Fig. 3. Error rate curve with different iterations

As we can see in Fig. 3, with the increasing number of iterations, the error rate is decreasing distinctly. What's more, the fourth method is a little better than the third one.

5 Conclusion

We developed a new color cast detection method which is based on the multi-feature extraction and shown that the method is universal to natural images. The experiments in the paper have shown that the method is pretty well with high detection rate and low error rate. So we can conclude that the method is suitable for the color cast detection before color correction and can help reduce the situation of over-correction. But because of the quantitative limitation of the standard color images, the experiment result is not as good as we think, in the next step, we can extend the numbers of the samples to get a better result.

Acknowledgment. The paper is sponsored by The National Key Technology R&D Program of the Ministry of Science and Technology, No.2013BAH03B01, The State Basic Research Development Program (“863 Program”), No.2012AA12A305 and Special Project on the Integration of Industry, Education and Research of Guangdong Province, No.2012A090300017.

References

1. Wang, H.: The Research of Color Cast Correction of Color Image. Jilin University, Changchun (2011)
2. Xu, X., Cai, Y., Liu, X., et al.: Improved Grey World Color Correction Algorithms. *Acta Photonica Sinica* 39(3), 559–564 (2010)
3. Zheng, J., Hao, C., Lei, F., et al.: Automatic Illuminations Detection and Color Correction of Image Using Chromatic Histogram Characters. *Journal of Image and Graphics* 8(9), 1001–1007 (2003)
4. Li, F., Jin, H.: Approach to detect digital image color cast based on image analysis. *Journal of Jiangsu University (natural science edition)* 25(5), 430–433 (2004)
5. Hu, B., Lin, Q., Chen, G., et al.: Automatic White Balance Based on Prior Information. *Journal of Circuits and Systems* 6(2), 25–28 (2001)
6. Du, X.: Research on Several Image Processing Problems as the Simulation of Vision Mechanisms. University of Electronic Science & Technology of China, Chengdu (2012)
7. Meylan, L.: High Dynamic Range Image Rendering with a Retinex-based Adaptive Filter. *IEEE Transactions on Image Processing* 15(9), 2820–2830 (2006)
8. Wang, X., Lu, Y., Song, S., et al.: Face recognition based on Gabor wavelet transform and modular PCA. *Computer Engineering and Applications* 48(3), 176–178 (2012)
9. Zhang, G., Ma, Z.: An Approach of Using Gabor Wavelets for Texture Feature Extraction. *Journal of Image and Graphics* 15(2), 247–254 (2010)
10. Ho, W.T., Tay, Y.H.: On Detecting Spatially Similar and Dissimilar Objects using AdaBoost. In: *International Symposium on Information Technology, ITSIM 2008*, pp. 1–5. IEEE (2008)
11. Li, C., Ding, X., Wu, Y.: An Algorithm for Text Location in Images Based on Histogram Features and AdaBoost. *Journal of Image and Graphics* 11(3), 325–331 (2006)
12. Ciurea, F., Funt, B.: A Large Image Database for Color Constancy Research. In: *Proc. IS&T/SID’s Color Imaging Conference*, pp. 160–164. The SunBurst Resort, Scottsdale (2004)

Research on an Extracting Method to Assist Manual Identification of Targets

Yue Shi and Guoying Zhang

School of Mechanical Electronic and Information Engineering, China University of Mining and Technology, Beijing, China
{shiyue1989,zhangguoying1101}@163.com

Abstract. The interpretation of the target recognition in remote sensing image is normally manually implemented by interpreters. In this essay, we propose to analyze the method for rapid extraction of suspected targets (method that can be used to assist manual interpretation) was proposed. The method is based on the image gray-scale characteristics. First of all, resolution reduction and enhancement should be used to preprocess the image. Then, through a series of means, including binarization, erosion, big target extraction and dilation, suspected targets are extracted. The experimental results show that the method can effectively and quickly extract targets from the remote sensing image which owns the features that the gray value of background area is mussy and the object area is homogeneous.

Keywords: Remote sensing image, image processing and target extraction.

1 Introduction

The traditional ways to extract targets from remote sensing image mainly rely on the manual operation of photographic interpreter. But manual methods have many defects such as big workload and low efficiency. Furthermore, affected by the weather condition, quality of the camera equipments or visual angle, etc., the objects in remote sensing image are different from the objects in the actual scene, which requires that the photographic interpreter has a lot of skills and experience. Thus, training a professional photographic interpreter takes a lot of time and energy.

Since the 1970's, many countries around the world have heavily invested to the study and the development of computer aided image interpretation system. Its workflow is as the following: firstly, the image is processed by computer to exclude the no value area and mark suspected objects. And then, the result image of the last step is interpreted by photographic interpreter. The application of computer image processing technology in remote sensing field greatly reduces the workload and improves the working speed and the precision of the interpretation. With the rapid development of the satellite remote sensing technology, satellites, the photos of which have the resolution less than 1 mile, keep emerging. Those high resolution images provide an objective foundation for object extraction method that is based on the gray level characteristics.

In this research paper, we introduce a method which allows the use of the remote sensing image’s gray-scale characteristic to extract quickly the suspected targets. The method aims to reduce human consumption in the process of interpretation and, at the same time, alleviate the pressure of the photographic interpreter training.

The rest of this essay is arranged in the following way: the second part introduces image pretreatment; the third part introduces the suspected target extraction method; the fourth part gives the experimental and analysis results; and finally, the fifth part is the conclusion.

2 Image Pretreatment

2.1 Reducing Image Resolution in One Process

Because the remote sensing image has high resolution, it requires not only a huge amount of storage, but also a reduced speed processing. Reducing the resolution of the processed image can help to solve above mentioned problems. For this purpose, two solutions are put forward in the following part.

- (1) Resolution Reduction

If the size of the original image $f(a, b)$ is $A \times B$, compressing the original image to compress the image $g(a_N, b_N)$, which size is $A_N \times B_N$, proportionally. N , equals

to $\frac{A}{A_N}$ (also means $\frac{B}{B_N}$), is the size of the compression window. Reducing pixels in

the window to one pixel point, which gray value $g(x_N, y_N)$ is the average value of those pixels in the window.

$$g(x_N, y_N) = \frac{\sum_{(x,y) \in M} f(x, y)}{N * N} \tag{1}$$

N is the size of the template. M is the set of pixels in the template. The original image’s pixel coordinate (x, y) corresponding to the compressed image’s pixel coordi-

$$\text{nate } (x_N, y_N) = \left(\frac{x}{N}, \frac{y}{N}\right).$$

Since the image compression reduces the size of the original image, the amount of image data is naturally decreased; and consequently this contributes to consume minor storage capacity and improve the processing speed.

- (2) Image Segmentation

N , the number of blocks after the segmentation, is determined by the size of image. In order to ensure that one extracted target exists completely in at least one image after

image segmentation, those segmented images should have some parts of the pixels to overlap. The size of the overlap area requires twice as much as the size of the biggest extracted target.

The size of the original image is $A \times B$ and the size of the biggest extracted target is not bigger than $bSize \times bSize$. If the original image is segmented into 4 parts, the size of images after the segmentation would be $(\frac{A}{2} + bSize) \times (\frac{B}{2} + bSize)$. Parallel processing is used to process those segmentation images in order to improve processing speed.

2.2 Image Enhancement

The purpose of the image enhancement is to highlight useful information of the image and enlarge the differences between different objects. It makes a good foundation for the following steps. The simple image enhancement method, which is commonly used, is the linear gray level transformation, histogram equalization, etc.

The linear gray level transformation is based on linear formula. Suppose that the gray-scale range of image $f(a, b)$ is $[f_{\min}, f_{\max}]$, the gray-scale range of output image $g(a, b)$ extends to $[0, 255]$. The transformation formula is:

$$g(x, y) = \frac{f(x, y) - f_{\min}}{f_{\max} - f_{\min}} \times 255 \quad (2)$$

The histogram equalization is based on a statistical theory. We suppose that an image has n pixels and l different gray levels. The original image's gray-scale is represented by r . The gray-scale of image, which is processed by histogram equalization algorithm, is s . n_k equals to the number of pixels which gray-scale is r_k , and then the appearance frequency of the k th gray-scale is expressed as:

$$P_r(r_k) = \frac{n_k}{n} \quad (3)$$

In the expression, $0 \leq r_k \leq 1$, $k = 0, 1, \dots, l-1$

After the equalization, each pixel's gray value is:

$$s_k = \sum_{j=0}^k P_r(r_j) \quad (4)$$

In the expression, $0 \leq r_j \leq 1$, $k = 0, 1, \dots, l-1$

3 Suspected Target Extraction

Targets mostly hide in the lawn or forest. In remote sensing image, the distribution of the gray value in those regions is uneven, but the gray value distribution of target area is relatively homogeneous. This feature can be used to remove the useless background. Furthermore, the number of pixels in the area of suspected target is much larger than the number of pixels in other useless objects' area; for this reason, those areas can be removed along with background.

3.1 Binarization

Binarization is defined as a means to transform the color information into binary image. Assume that the threshold value is *Threshold*, the input image is $f(x, y)$ and the output image is $g(x, y)$; therefore the binarization can be expressed by the following formula:

$$g(x, y) = \begin{cases} 1 & f(x, y) \geq \textit{Threshold} \\ 0 & f(x, y) < \textit{Threshold} \end{cases} \quad (5)$$

The key function of the binarization is to select an appropriate threshold since a good threshold can reserve useful image information and eliminate the distracting information as much as possible. The fixed threshold method and dynamic threshold method are two common ways to obtain threshold. The fixed threshold method uses the mean value of all pixels in the image. Although this method is simple and the execution speed is fast, the mistake rate is rather high, since it only considers the overall image and ignores the local information. Here is a kind of dynamic threshold method based on double window.

We suppose that the input image is $f(x, y)$ and the output image is $g(x, y)$. We set that two windows' size respectively are W_1 、 W_2 ($W_1 > W_2$). For a pixel $f(a, b)$, the average gray value of the pixels for the two windows, the central point of which is $f(a, b)$, are Ave_1 and Ave_2 . The smaller value is thus considered as the threshold. The formula is as the following:

$$Ave_1 = \frac{\sum_{(x,y) \in W_1} f(x, y)}{W_1 \times W_1} \quad (6)$$

$$Ave_2 = \frac{\sum_{(x,y) \in W_2} f(x, y)}{W_2 \times W_2} \quad (7)$$

$$\text{Threshold} = \min(\text{Ave}_1, \text{Ave}_2) \quad (8)$$

3.2 Erosion

b is the structural element, that is the template of erosion. f is the set of images. The formula of erosion is as the following:

$$f \otimes b = \left\{ x \mid (\hat{b})_x \subseteq f \right\} \quad (9)$$

The formula shows the erosion means that any element in set x can be the center of template b (note that b is still included in f).

3.3 Target Extraction

The target extraction algorithm calculates the number of pixels in each white area and sorts in descending order. Only the top N areas' pixels remain the same and pixels in other areas are set to zero. The algorithm is described as the following:

(1) The searching window's size $Size$ and N (the number of areas which are reserved) are given. The setting mark image $flag[nHeight][nWidth]$ ($nHeight$ and $nWidth$ are respectively the length and the width of the original image), queue que and counting array $area[1000]$. All the pixels in the mark image and counting array $area[1000]$ are initialized to zero.

(2) The binarization image is scanned line by line until the first pixel (the gray value of which is not zero). If this pixel's corresponding point in the mark image is 0, this corresponding point's gray value would be set to n (n is the area numbering from 1 to 1000) and its coordinate is put into the rear of que .

(3) In the binarization image, the coordinate in the head node of the queue que is used as the center point of searching window. The pixels, in the searching window, are find, the gray value of which is not zero and its corresponding point, in the mark image, is unmarked. Those pixels' coordinate are put in the rear of que separately and its corresponding point's gray value, in the mark image, is set to n .

(4) Repeating step (3) until que is empty;

(5) Executing Step (2) (3) (4), until completion of image scanning;

(6) Counting the number of pixels in each area and putting the number in $area[1000]$ ($area[n]$ holds the number of pixels in the area n).

(7) Sorting $area[1000]$ in descending order. Only reserving the top N areas and setting pixels, in the other areas, to zero.

3.4 Dilation

b is the structural element, that is the template of dilation. f is the set of images. The formula of dilation is as the following:

$$f \oplus b = \left\{ x \mid \left[(\hat{b})_x \cap f \neq \phi \right] \right\} \quad (10)$$

In the expression, \hat{b} is the mapping of b . \hat{b} and b are symmetric about origin.

4 Experiment Process and Analysis

4.1 Image Pretreatment

Fig. 1 and Fig. 2 are the results of two processes, which are respectively resolution reduction and image segmentation. The advantage of the image segmentation method is retaining all the pixels, which contributes to high processing precision, and the processing speed is also improved. But when using this method, the rough size of the target should be known in advance. Besides, the problem that is taking up too much processing memory is not solved. Although the processing speed is improved, parallel processing increases the cost of execution. For the larger target, resolution reduction is better. Since it increases the processing speed, it preserves well the target's contour. The proposed approach, in this essay, aims to improve the speed of target extraction. Hence the resolution reduction method is adopted and its result is used for further processing.

After the resolution reduction, the overall image becomes gloomy and image's details become blurring. In order to facilitate further processing, the improving of the visual effect of image is needed.

Fig. 3 and Fig. 4 are the results which are treated respectively with the linear gray level transformation as well as the histogram equalization method. After being processed by the means of the histogram equalization, the contrast of the image is significantly enhanced, but the top right corner is over-bright and blurring. Comparison shows that the processing result of the linear gray level transformation is better because the visual effect of image is effectively improved. Besides, the computational complexity of the linear gray level transformation is smaller and the processing speed is faster. In order to achieve the goal, the objects can be extracted quickly and exactly because the linear gray level transformation method had been adopted.

4.2 Suspected Target Extraction

The binarization result of the double window binarization method depends on the size of the two windows. According to the experiment, choosing 25 and 50 as the size of two windows respectively can obtain a relatively better effect and the processing speed is faster. Fig. 5 is the result image of the double window binarization method.



Fig. 1. Resolution Reduction



Fig. 2. Image Segmentation



Fig. 3. Linear Gray Level Transformation



Fig. 4. Histogram Equalization

From the Fig. 5, one feature can be found out: that is objects are white inside basically and connect into blocks while the pixels in background area are randomly black or white. The white point in the background area is scattered and can not be linked up into larger blocks. Erosion is used to decrease the connection degree of white points in the background area further. The erosion template, the size of which is 3×3 , is used to process image. The result is Fig. 6.

In Fig. 6, the area of lawn is similar to salt-and-pepper noise and the suspected targets occupy the large white blocks. The above characteristic can be used to eliminate useless areas and reserve suspected targets. On this picture, take 10 as N , which is the number of blocks to be reserved.

After the observation of Fig. 7, one can realize that suspected targets have been extracted out. But there is some lack of pixels and outline is not obvious. The dilation processing can make the white area expanded and diminish the hole inside the target. So, template 3×3 is used to process the image and Fig. 8 is the result.



Fig. 5. Double Window Binarization

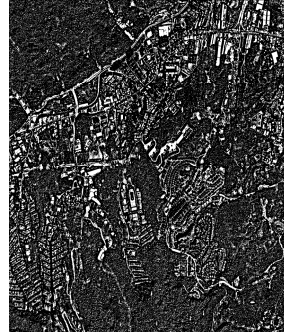


Fig. 6. Erosion

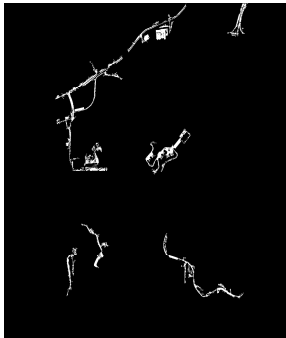


Fig. 7. Extracted Target

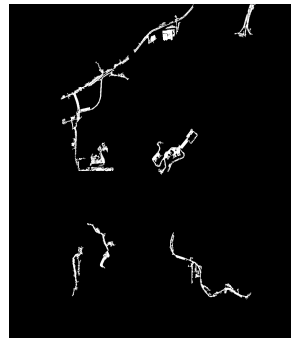


Fig. 8. Dilation

Though the observation of Fig. 8, one can find that the useless information is removed and the suspected targets have been extracted. Photographic interpreter can find out targets in the remaining suspected targets. The workload is greatly reduced and the interpretation precision is improved.

4.3 Applicability Analysis

In order to verify the applicability of this method, we did experiments on a series of pictures. According to the actual operation, some of the parameters used in the experiment are adjusted.

The experimental results show that the method can achieve rapid and effective extraction through the remote sensing image owns complex background and objects area with homogeneous gray value. The missing rate tends to be zero. Especially, for the linear targets, such as airport runways and roads, it works the best (as shown in Fig. 9 and Fig. 10). For targets, which have non-uniform inside, miss rate is high (as shown in Fig. 11).

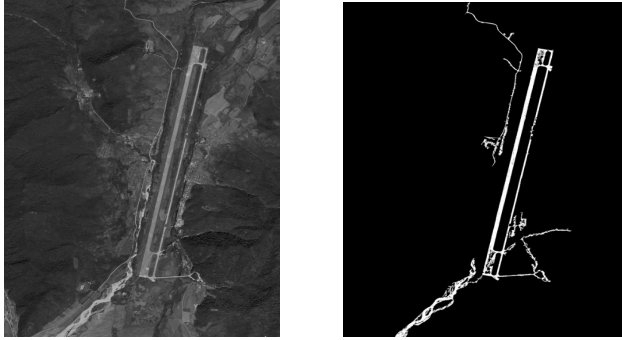


Fig. 9. Experiment1 (success)

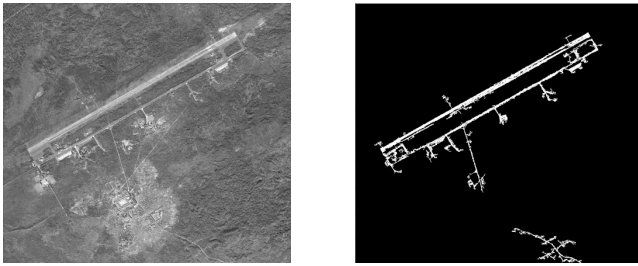


Fig. 10. Experiment2 (success)

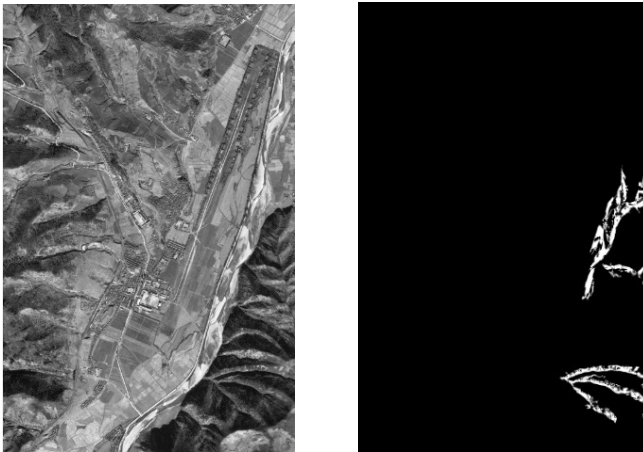


Fig. 11. Experiment3 (failure)

5 Conclusion

In this research paper, we put forward a method of rapid extraction for suspected target which can be used to assist manual interpretation. By comparing the results of different pretreatment method, we select the resolution reduction and the linear gray

level transformation to preprocess image. We will also explain how the double window binarization, erosion, suspected target extraction and dilation are used to remove useless image information in order to eventually extract the suspected target. This method can be used to help photographic interpreter for further interpretation.

The experimental results show that the method can achieve rapid and effective extraction through the remote sensing image which has complex background and objects area with homogeneous gray value. The missing rate tends to zero. The implementation of target extraction, the internal of which is non-uniform, represents the content of further study.

References

1. Yang, Y., Zhao, R., Wang, W.: Automatic Building Detection in Aerial Image. *Computer Engineering* 28(8), 20–21, 27 (2002)
2. Qin, Q., Yuan, Y., Lu, R.: The Recognition of Various Types of Water Bodies on Satellite Image. *Geo Graphical Research* 20(1), 62–67 (2001)
3. Gong, S., Liu, C., Wang, Q.: *Digital Image Processing and Analysis*, pp. 48-51, 177-184, 199-200. Tsinghua University Press (2006)
4. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing (the Third Edition)*, pp. 68–77, 402-406. Electronic Industry Press (2011)

A Face Recognition under Varying Illumination

Haodong Song^{1,2}, Xiaozhu Lin¹, and Zhanlin Liu^{1,2}

¹ School of Information Engineer, Beijing Institute of Petro-chemical Technology,
Beijing 102617, China

{hd_biptgra, linxiao, zhanlinliu}@bipt.edu.cn

² College of Information Science and Technology,
Beijing University of Chemical Technology, Beijing 100029, China

Abstract. Face recognition is one of the focus studies in biometrics technology. The recognition accuracy always changes drastically in different environment, especially when it is affected by the illumination. Retinex is a method which utilizes illumination invariant, but it ignores contributions of low frequency component to face recognition. In this research paper, we propose a face recognition method based on retinex and wavelet transformation. First, illumination invariant and variant are generated by the retinex theory. Second, decompose the illumination component via wavelet transformation and set its low-frequency coefficients to zero. In doing so, the processed illumination component is obtained by inverting the transformation. In the end, a new image is acquired by restructuring the two components. The recognition experiment will demonstrate that the proposed method ensures good performance in illumination environment.

Keywords: Face recognition, Retinex, illumination invariant and wavelet transformation.

1 Introduction

Face recognition is a challenging field in pattern recognition. It is difficult to satisfy the practice application because of the low recognition accuracy under illumination, pose and facial expression. In order to solve the problem of illumination, researchers proposed a variety of approaches, such as illumination normalization [1,2], 3D illumination model [3,4] and illumination invariant [5,6]. The illumination normalization is a method based on image gray transformation. Though it can be easily implemented, the result doesn't satisfy the demand under complex illumination. The illumination model, a way establish face model by different images under different illumination conditions, achieves a good accuracy in laboratory. However, it's impossible to obtain a plenty of images in real time face recognition system. For the last category, Land [6] proposed a retinex theory which suggests that the color of the object depends on the capability of the light wave reflection and is consistent without the effect of light discontinuity.

As mentioned above, we can extract reflectance from original image to recognize. This theory also can be formulated by equation (1):

$$I(x, y) = R(x, y) \cdot L(x, y) \quad (1)$$

Here, I represents original image; R represents reflectance and L represents illumination. We make a logarithmic transformation of equation (1), it is showed by equation (2):

$$\log(R(x, y)) = \log(I(x, y)) - \log(L(x, y)) \quad (2)$$

If we can estimate the illumination, the reflectance can be calculated as equation (2). A common assumption is that illumination spatially changes slowly, the low frequency of image serves as illumination L . Thus, some smoothing filters were proposed to obtain L . Park et al [7] gave a retinex method based on adaptive smoothing which could effectively extract the light discontinuity from image under illumination.

As mentioned above, retinex is a means to utilize illumination invariant and discard the low frequency (illumination L). But the low frequency contains some basic information of human face. Therefore we want to apply part of low frequency information to improve the recognition accuracy. In this essay, we propose a novel face recognition based on retinex and wavelet transformation. First, we decompose the image into reflectance and illumination by retinex. Second, eyes positions which are located by the eye detection from reflectance are employed to align and segment the face in order to identify an effective recognition area. Third, the illumination is decomposed by the wavelet transformation and its low-frequency coefficients are set to zero; then, the processed illumination image is obtained by using the inverse transformation. At last, the two components compose a new image. The essay is organized following the following way: in section 2, a face recognition method using adaptive smoothing retinex is reviewed and its shortcoming will be analyzed; in section 3, we introduce the wavelet transformation and present our method in details; in section 4, the result of experiment is presented; and finally, we conclude this essay in section 5.

2 Face Recognition Based on Retinex

2.1 Adaptive Smoothing Retinex

Adaptive smoothing retinex [7] is an algorithm which iteratively convolves the original image with an averaging mask whose coefficients reflect the discontinuity level of the original image at each point. Its main equations can be showed as in equation (3) to (5)

$$L^{(t+1)}(x, y) = \frac{1}{N^{(t)}(x, y)} \sum_{i=-1}^1 \sum_{j=-1}^1 L^{(t)}(x+i, y+j) \cdot w^{(t)}(x+i, y+j) \quad (3)$$

$$N^{(t)}(x, y) = \sum_{i=-1}^1 \sum_{j=-1}^1 w^{(t)}(x+i, y+j) \quad (4)$$

$$\begin{cases} w(x, y) = \alpha(x, y) \cdot \beta(x, y) \\ \alpha(x, y) = g(\tau(x, y), h) \\ \beta(x, y) = g(|\nabla I(x, y)|, S) \end{cases} \quad (5)$$

In equation (3), L represents illumination and w represents mask. The w consists of two masks which can regulate its weight at each point by calculating the light discontinuity. So the $L^{(t+1)}(x, y)$ may contain more light in once iteration. In addition, the $(t+1)$ th L may involve less light than in (t) th with iteration counting. Consequently, R may include more light shadow for less light in L , which can be expressed by (2). In conclude, reflectance will become worse with the iteration increasing.

2.2 Face Processing

There are two steps in the processing: first, the retinex should be applied to obtain invariant; and second, an eye detection is used to align and segment human face in the reflectance. For the first one, the detail has been expressed in [7]; and in the second case, there are always useless backgrounds in the image. Besides, the face may tilt when people is under an uncontrolled environment. Therefore, many researchers will manually align and segment the images at first. For this reason, we are trying to designs an auto face processing method in this research paper. We use algorithms [8,9] to locate eyes positions and use a line to connect the two eyes. Then, we select the center of the line as reference point and extract an effective square with side length of $1.8l$ in the reflectance (l is the distance between two eyes. Moreover, the distances of reference point to left and right side are $0.9l$ and $0.9l$. $0.5l$ and $1.3l$ are the distances to top and bottom side). The processed face images are illustrated in Fig.1.

As showed in Fig.1 (a), the original image is affected by the illumination and the background, besides, the face slightly tilt to left. Fig.1 (b) to Fig.1 (d) has the some results acquired by the method presented above. Obviously, there are not only an effective field obtained by aligning and segment, but also less illumination left. After that, the Fig.1 (d) will be used to recognize.

2.3 The Effect of Illumination Element to Face Recognition

According to section 2.1, the quality of invariant is related to iteration t . When t is small, the shadow is removed effectively in the invariant. When t is lager, the reverse would be true. So we suppose that a satisfying recognition accuracy should be obtained with a small t . However, a series of experiments suggest that this opinion is not true. We select 10 different people with 10 illuminations for each one. Then, the images are divided into two sets, one and half are used to train and others are used to test. All images are processed with measure in section 2.2. In the experiment, t increases with 10 and the classifier is the NN (Nearest Neighbor).

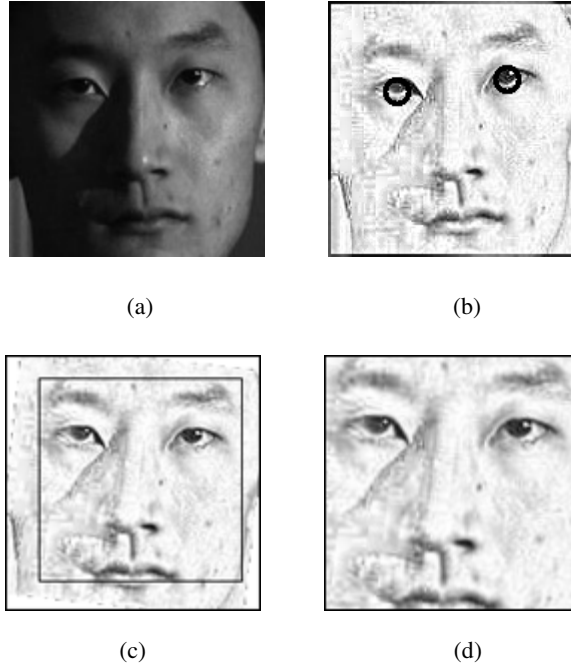
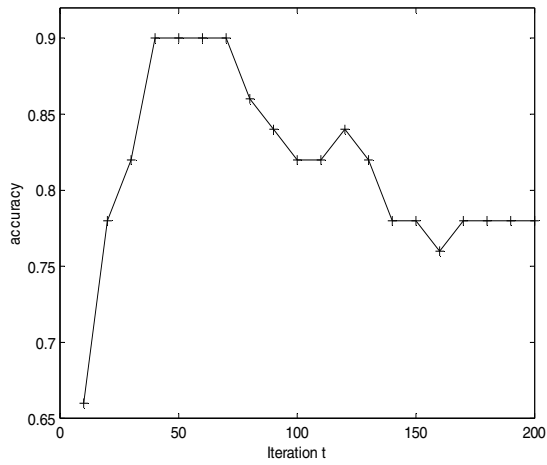
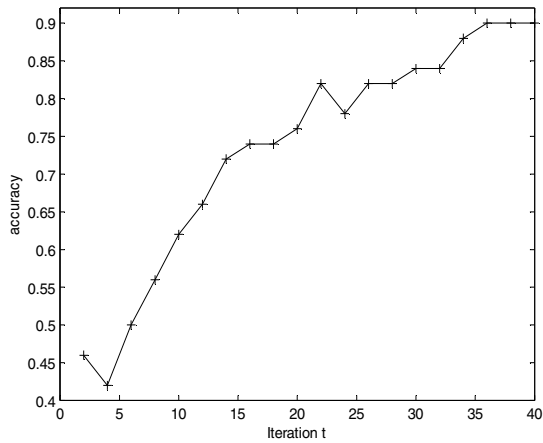


Fig. 1. The results of processed image. (a) Original image. (b) Eye detection. (c) Effective field (d) Standardization

Fig.2 illustrates the effect of iteration to recognition. In Fig.2 (a), when $t < 70$, the accuracy increases with t increasing, while $t > 70$, the accuracy decreases because the invariant is affected by more shadow that develops gradually. In order to clearly analyze the small t , we make a new experiment in which we apply t from 2 to 40 with 2 increasing. The result is showed in Fig.2 (b). Though we infer high accuracy that should be found with small t , unexpectedly, the experiment demonstrates reverse result. Low accuracy reveals that it is not reasonable to discard all illumination to raise the accuracy. Furthermore, when t increases, the accuracy improves well because some low frequency information are preserved in the reflectance. Xie et al [10] pointed out that the low frequency contains basic information of face which may offer additional content to distinguish different people in their research on the effect of large- and small scale features to face recognition. Whereas the retinex only uses high frequency and remove all low frequency, which may lose important information. So we believe that some information should be extracted to join in the invariant from low frequency (illumination) to improve face recognition.



(a)



(b)

Fig. 2. The effect of iteration t to recognition. (a) Iteration t from 10 to 200. (b) Iteration t from 2 to 40

3 Improvement Using Wavelet Transform

The wavelet transformation [11] is a time-frequency analysis based on short-time fourier transformation. One of features is Multi-resolution analysis. This method can decompose a signal in coarse-to-fine to acquire some key information by selecting different scales factors of scale function and wavelet function. For image, we usually

apply dimensional wavelet transformation to decompose and reconstruct it with the Mallat algorithm. First, we design a low-pass and high-pass filter according to the kinds of wavelets. Then, we use two filter to make the convolutions with image. This way, the image is divided into four sub-bands as Fig.3 shows.

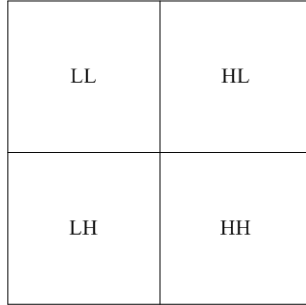


Fig. 3. Dimensional wavelet transform of image

Fig. 3 illustrates the result of the wavelet transformation at once. *LL* represents the low frequency of sub-band which contains the basic information of an image, while *HL*, *LH* and *HH* represent three high frequency sub-bands. If the wavelets transformation is executed to the *LL*, it can be decomposed into other four sub-bands with smaller scales. Continue to decompose *LL* of each sub-band will provide more sub-bands in multi-resolutions. Thus, we can obtain some information in different frequency by wavelet transformation.

With the above analysis, we are applying the wavelet transformation to decompose the illumination acquired via the retinex in this essay. Then, we will set the low frequency coefficients to zeros to obtain a new “illumination” by inverse transformation. The new “illumination” can be seen as “middle frequency” in original image. At last, invariant and the “illumination” are used to construct a face for recognizing. Here is the detailed explanation of our method:

- 1) Utilize the retinex to decompose the original image into illumination invariant and illumination (low frequency component).
- 2) Locate eyes in reflectance via the approaches in [5,6], then rotate and segment the reflectance and illumination to acquire the effective fields according to the eye position.
- 3) Decompose the illumination by the wavelet transformation and set $LL(i)$ coefficients to zeros, then make the inverse transformation to acquire new “illumination”, here the i means the i th layer decomposition.
- 4) Reconstruct new image according to equation (2).

The process is illustrated in Fig.4

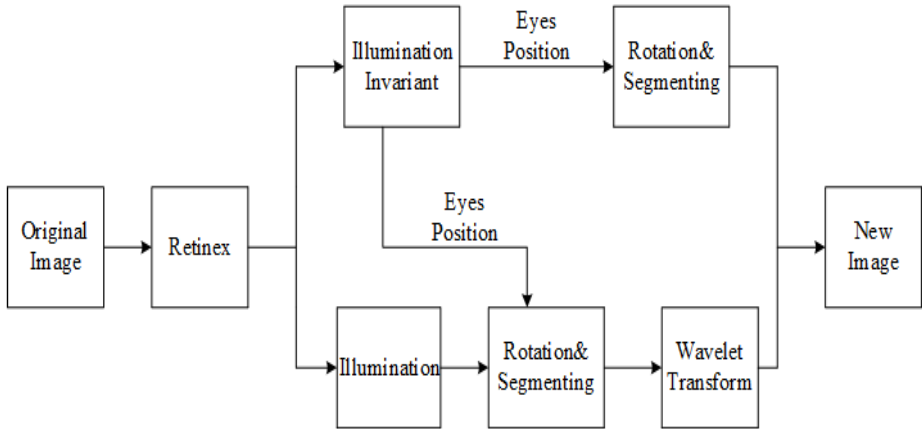


Fig. 4. Diagram of proposed method

4 Experiment and Analysis

To verify our method, we have designed an experiment with CMU-PIE. The CMU-PIE contains a plenty of face images in many complex conditions, such as pose, illumination and expression. Since we only consider the effect of illumination to face recognition, we chose 100 front faces of 10 individuals in 10 different illumination angles from the face database. We compared our method and the adaptive smoothing retinex. For parameters in the experiment, we chose 40 for the iteration t and 2 layers for wavelet transformation. All processed images are resized to 100×100 .



Fig. 5. Examples of two methods

Fig. 5 illustrates the processed face in two methods. The first line is the original images. The second line and third line are respectively processed by the adaptive smoothing retinex and our method. We realized that the original images are affected by

the illumination from different angles, moreover, the background may cause error recognition. In contrast, the processed images, regardless of the retinex or our method, can solve the problem of illumination and useless background. Compare the 2th and 3th lines, the images in 2th line are much white than those in 3th line. Though distinct edge isn't affected by illumination, it can't present enough important basic information which distinguishes some people with similar edge. Consequently, our method adds some low frequency component to invariant to add additional information to improve accuracy.

Table 1 presents the recognition accuracies of three conditions. In order to verify the effectiveness of our method, we employ the recognition rate of original condition as a standard. Obviously, the accuracy of our method is higher than that of the standard in table 1. Compare our method and the retinex, the accuracy of our method is 4% higher than that using retinex. The reason why our method is better is that an appropriate low frequency component adds useful information to reflectance to make the images distinguishing easy.

Table 1. Comparison of different methods

Methods	Accuracy
Original image	82%
Retinex	90%
Our method	94%

5 Conclusion

Illumination, pose and expression are the three important factors to cause low recognition accuracy in face recognition. This essay proposes a method based on retinex and the wavelet transformation to the problem of face recognition under the illumination condition. First, the illumination invariant and variant are generated by retinex theory. Second, the reflectance and illumination are rotated and segmented to identify the effective fields according to eye positions. Third, decompose the illumination component via the wavelet transformation and set its low-frequency coefficients to zero. Therefore, the processed illumination component is obtained by the inverse transformation. Last, restructure the two components to a new image. The experiment demonstrates that our method performs well. In the future, we will prepare to employ this method to the image acquired from camera in real time, this will make our method well applicable to the public environment.

References

1. Wei, C.C., Wang, X.P., Yan, J.W., et al.: Method for eliminating the Illumination Effects of Face Recognition (一种消除光照影响的人脸识别方法). *Electronic Test* 7, 19-23 (2012)

2. Lee, P.H., Wu, S.W., Hung, Y.P.: Illumination Compensation Using Oriented Local Histogram Equalization and Its Application to Face Recognition. *IEEE Transaction on Image Processing* 9, 4280–4289 (2012)
3. Georghiades, A.S., Belhumeur, P.N., Riegman, D.K.: From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *IEEE Transaction on Pattern on Pattern Analysis and Machine Intelligent* 6, 643–660 (2002)
4. Li, S.M., Liu, D.J., Sun, S.X., et al.: Research on Face Recognition Based on Illumination Subspace (基于光照子空间的人脸识别研究). *Journal of Chengdu University of Technology (Science & Technology Edition)* 5, 588–592 (2011)
5. Fan, C.N., Zhang, F.Y.: Illumination Invariant Extracting Algorithm Based on Nonsub-sampled Contourlet Transform (基于非下采样Contourlet变换的光照不变量提取算法). *Signal Processing* 4, 507–513 (2012)
6. Land, E.H.: An Alternative Technique for Computation of the Designator in the Retinex Theory of Color Vision. *Proceeding of National Academy of Science* 10, 3078–3080 (1986)
7. Park, Y.K., Park, S.L., Kim, J.K.: Retinex Method Based on Adaptive Smoothing for Illumination Invariant Face Recognition. *Signal Process* 8, 1929–1945 (2008)
8. Jung, C., Sun, T., Jiao, L.C.: Eye Detection under Varying Illumination Using the Retinex Theory. *Neurocomputing* 113, 130–137 (2013)
9. Song, H.D., Lin, X.Z.: Eye Location Under Varying Illumination (变化光照条件下人眼定位方法). *Journal of Beijing Institute of Graphic Communication* 2, 62–67 (2014)
10. Xie, X.H., Zheng, W.S., Lai, J.H., et al.: Normalization of Face Illumination Based on Large- and Small-scale Features. *IEEE Transaction on Image Processing* 7, 1807–1821 (2011)
11. Sun, Y.K.: *Wavelet Transform and Processing Technology of Image and Graph (小波变换与图像、图形处理技术)*. Tsinghua University Press, Beijing (2012)

CS-FREAK: An Improved Binary Descriptor

Jianyong Wang, Xuemei Wang, Xiaogang Yang, and Aigang Zhao

Department of Automation, Xi'an Institution of High-Tech
Xi'an, China
TinkerSpy@163.com

Abstract. A large number of vision applications rely on matching key points across images, its main problem is to find a fast and robust key point descriptor and a matching strategy. This paper presents a two-step matching strategy based on voting and an improved binary descriptor CS-FREAK by adding the neighborhood intensity information of the sampling points to the FREAK descriptor. This method divides the matching task into two steps, firstly simplify the FREAK[1] 8-layer retina model to a 5-layer one and construct a binary descriptor, secondly encode the neighborhood intensity information of the center symmetry sampling points, and then create a 16-dimensional histogram according to a pre-constructed index table, which is the basis for voting strategy. This two-step matching strategy can improve learning efficiency meanwhile enhance the descriptor identification ability, and improve the matching accuracy. Experimental results show that the accuracy of the matching method is superior to SIFT and FREAK.

Keywords: Point matching, binary descriptor, two-step matching strategy, FREAK.

1 Introduction

Image matching plays a key role in computer vision, image stitching, target recognition and other fields [3][4]. Image matching consists of three steps: feature points detect, feature points description and matching.

After obtaining the feature points, adding an appropriate description is a critical work, which greatly affects the subsequent efficiency of image matching. Lowe DG [2] et al. proved that SIFT descriptor is more stable through conducting a performance evaluation experiments, in which the images were processed by changing blur, light, scale and a certain perspective transformation. The experiment results show that SIFT descriptor can get better matching results compared with other descriptors including shape context information, complex filtering, invariant moments [5], etc. But there exist shortcomings: the descriptors are made of 128-dimensional vector which is too high, a large number of feature points got involved in matching, and the search measurements are relatively time-consuming. Therefore, the subsequent emergence of a variety of descriptors are improvements of SIFT descriptors, such as the PCA-SIFT using principal component analysis to reduce the dimensionality [6], GLOH descriptors using the log-polar grid interval instead of grid interval [7], SURF descriptor

accelerating the process by introducing integrating graphics into its process of describing [8]. But dimensions of these descriptors remain high, which are unsatisfactory in terms of real-time practice, therefore the binary descriptors become hotspot recently. A clear advantage of binary descriptors is that the Hamming distance (bitwise XOR followed by a bit count) can replace the usual Euclidean distance, eliminating the common matching strategy such as building a K-d tree. Calonder et al. put forward the BRIEF[9] which is obtained by comparing the intensity of 512 pairs of pixels after applying a Gaussian smoothing to reduce the noise sensitivity; Ruble et al. improved the traditional FAST by adding orientation information and proposed the Oriented Fast and Rotated BRIEF (ORB)[11], which gets strong robustness to noise and rotation, which is obtained by comparing the intensity of random pixels pairs; Leutenegger et al. put forward BRISK descriptor[10] which is invariant to scale and rotation, their BRISK is obtained by comparing a limited number of points in a specific sampling pattern; Alahi et al, heuristically proposed FREAK[1] descriptor according to human retina system, that a cascade of binary strings is computed by efficiently comparing image intensities over a retinal sampling pattern.

In this paper, we propose a new method for feature description and a novel matching strategy based on FREAK descriptor. We simplified the 8-layer circle model to a 5-layer model which shortcut the description time, and added the neighborhood information of the fixed sampling points as the vote data to ensure the matching accuracy.

1.1 Two-Step Matching Strategy

FREAK descriptor is a binary bit string descriptor by thresholding the difference between pairs of receptive fields with their corresponding Gaussian kernel, which is called a binary test. FREAK takes a 8-layer model consist of 43 sampling points, [12] pointed out that, compared with BRIEF, conducting the binary test with the utilization of fixed sampling point improved training efficiency, but fixed sampling pattern may reject the optimal point collection, Alexandre Alahi also pointed out that FREAK is inspired by the human visual system and more precisely the retina [1], the focus of the study in terms of points selected is still important. [12] proposed a binary descriptor matching algorithm based on hierarchical learning method which combines the advantages of the fixed-point sampling mode and random sampling mode. proposed a fixed point of first use (3 layer 17 points) for training, and then point to where the circle from the candidate within the stratified random sampling for training learning model that combines the advantages of different sampling modes, thereby improving the learning efficiency. Drawing on the basis of this idea, we proposed a new method for feature description based on FREAK and a novel matching strategy based on voting.

1.2 Simplified Retina Modal

To utilize the neighborhood information of the key points more fully, we presents a two-step matching strategy, the first step is to use the simplified binary descriptors for matching, then the neighborhood information of the fixed sampling points is utilized to form a vote data, determining the final match result according to the number of

votes .The FREAK model is simplified to 5-layer with 25 sampling points, which can simplify the calculation and at the same time form a sufficient amount of the votes, as it shown in Figure 1,the detailed description of our matching strategy based on two-step voting is introduced as the following.

Extract the key point and simplify the FREAK descriptor. The simplified descriptor is generated according to the normal FREAK descriptor generation method, with the identification model proposed in [12], which is more reasonably discerning to quantify the differences between the test sequences.

$$D_a = \frac{\exp(-M^2(a))}{\arg \max R(a, b)}, s.t. 0 \leq a < N, \quad (1)$$

$$0 \leq b < N; a \neq b$$

where $M(a) = \frac{\sum_{0 \leq a < N} T(P_a)}{N} - 0.5$, $R(a, b)$ is the correlation between test sequence a and b . Sort the sampling pairs by D_a , and select the first N points binary test results (we take $N=64$ in the experiment) as the simplified binary descriptors. Use the hamming distance as similarity measure method, looking for their m nearest key points from images to be matched. This process obtained the crude matching results.

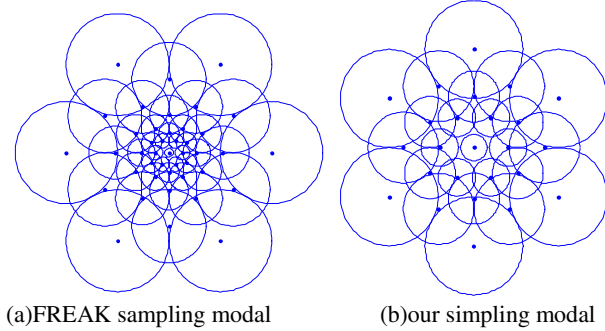


Fig. 1. The sampling modal

1.3 Center Symmetry FREAK and Voting Strategy

Utilize the neighborhood information of the sampling points as the vote data. We draw the ideas from [14] to encode the intensity information. In the 5-layer simplified model; there are 15 point pairs which are symmetry to the key point. As it shown in Figure 2, $P_{cs} = (P_1, P_2)$ is a symmetry pair, regard P_1 as the center, and connect P_1 and x_1 to get line₁, connect P_1 and x_2 to get line₂, connect P_1 and x_3 to get line₃, connect P_1 and x_4 to get line₄, then define a point set $\rho_j^1 = \{(x_j); x_j \in circle_i \cap x_m \in line_j\}$, $j=1,2,3,4$, similarly P_2 has $\rho_j^2 = \{(x'_j); x'_j \in circle_n \cap x'_m \in line_j\}$, $j=1,2,3,4$, define, vote_data =

$[T((x_1, x'_1)), T((x_2, x'_2)), T((x_3, x'_3)), T((x_4, x'_4))]$ $x_i \in \rho_j^1, x'_i \in \rho_j^2$, which is a four-bit binary description, there could be totally 16 kinds of values. Construct a index table, as is shown in [], find the corresponding position in the index table $index = bin2dec(vote_data) + 1$, where $bin2dec$ represents a conversion from binary to decimal. Define $\vartheta = (0, 0, 0, 1, \dots, 0)$, which is a 16-dimensional vector, count all the symmetry point pairs in ρ_j^1, ρ_j^2 to fill in the corresponding position of ϑ to construct a 16-dimensional histogram descriptor, which we call $CS - FREAK(P_1, P_2)$.

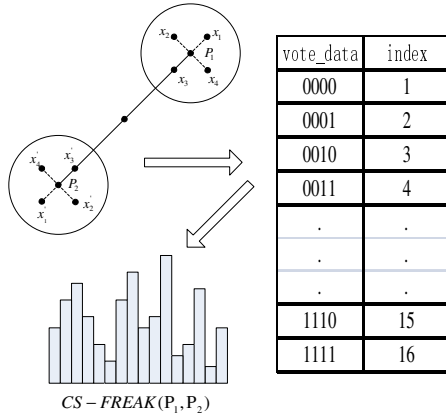


Fig. 2. The construction of the CS-FREAK descriptor

Based on the first step matching results, calculate the $CS - FREAK(P_1, P_2)$ Euclidean distance correlation between the key point in the training image and the corresponding one in the image to be matched, if the correlation is larger than a threshold, then take one vote, count votes, assume the point pairs with most votes as the correct matching result.

The specific steps to conduct the two-step matching are as following:

Step1. Extract multi-scale FAST key points;

Step2. Construct the simplified 5-layer model and build FREAK descriptor;

Step3. Conduct the rough match with the utilization of FREAK descriptor, ① if the hamming distance between the key point in the training image and a number of key points in the image to be matched is proximity, keep these point pairs and step into Step4; ② if the distance is large, which indicates the descriptor has strong identification force, correct match result is the point pair with minimum hamming distance;

Step4. Calculate the vote data of the candidate matching point pairs obtained in Step3;

Step5. According to the vote results, the correct matching result is the one with the most votes.

1.4 Orientation

In order to estimate the main direction of the key point, we mainly select pairs with symmetric receptive fields with respect to the center, and sum the estimated local gradients. As it shown in Figure 3, we select 33 points to compute the local gradients, Let G be the set of all the pairs used to compute the local gradients:

$$O = \frac{1}{M} \sum_{P_0 \in G} ((I(P_0^{r1}) - I(P_0^{r2}))) \frac{P_0^{r1} - P_0^{r2}}{\|P_0^{r1} - P_0^{r2}\|} \quad (2)$$

where M is the number of pairs in G and P_0^{ri} is the 2D vector of the spatial coordinates of the center of receptive field.

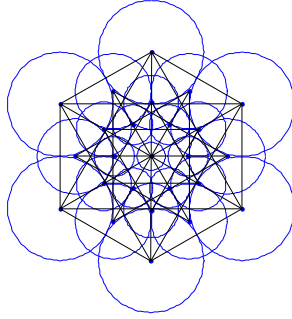


Fig 3. Illustration of the pairs selected to computed the orientation

2 Performance Evaluation

2.1 Dataset and Evaluation Criterion

To test the CS-FREAK descriptors performance, we build this new descriptor based on the OpenCV FREAK codes, and conduct a comparative experiment with SIFT and FREAK descriptors. We used the dataset introduced by Mikolajczyk and Schmid [7] as the test image data sets, to evaluate the rotation, change of scale, change of viewpoint, blur (Gaussian), and brightness change. All algorithms are running on Intel (R) Core (TM) i5-2.5 GHz, 2 G RAM with Window 7 + VS2010 + Opencv2.4.4. SIFT and FREAK algorithm codes are provided by OpenCV2.4.4 library. And we present our tests using the FAST detector also provided by OpenCV.

Descriptors are evaluated by means of precision-recall graphs as proposed in [13]. This criterion is based on the number of correct matches and the number of false matches obtained for an image pair. A match is correct if the overlap error < 0.5 .

$$recall = \frac{\#correct\ matches}{\#correspondences} \quad 1 - precision = \frac{\#false\ matches}{\#all\ matches} \quad (3)$$

where $\#correspondences$ is the ground truth number of matches.

2.2 Performance and Evaluation

As illustrated in Figure 4, CS_FREAK performs competitively with SIFT and FREAK, and these four tests all rank CS_FREAK as the robust to all the deformation, which can be interpreted as CS_FREAK inherits the advantages of FREAK, and also to be innovative in the usage of the identification model and the intensity information. Particularly CS_FREAK improves the performance on illumination changes (see Figure 4 (d)) due to the novel usage of intensity information. And because of the key point orientation, the CS-FREAK descriptor is invariant to rotation, so it also performs well on the rotation changes (see Figure 4 (c)).

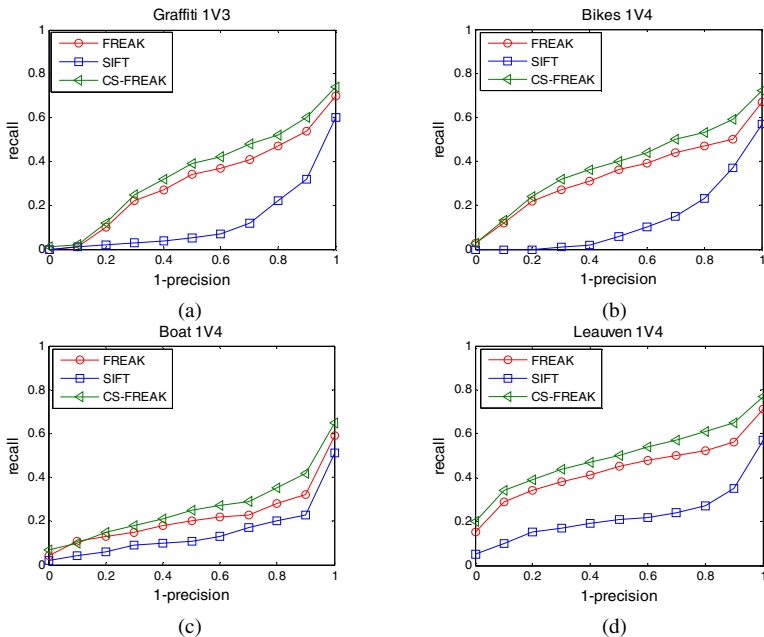


Fig. 4. Performance evaluation on the dataset

To investigate the time consuming of the algorithm, we calculate the time statistics of SIFT and FREAK, with a size of 800×640 in BMP Graffiti 1-3 statistical results, as shown in Table 1. Because of the simplified 5-layer modal, CS-FREAK is even faster than FREAK, though the voting strategy slows down the description and matching time theoretically, but not all the key points need to be conducted a voting step, only a part that have nearest hamming distance with several key points in the second image are involved, so CS-FREAK is faster than FREAK in description time, but not superior in the matching time, while calculating hamming distance takes less time than Euclidean distance.

Table 1. Computing time of SIFT, FREAK and CS_FREAK

	SIFT	FREAK	CS-FREAK
Points in first image	6003	6003	6003
Points in second image	7459	7459	7459
Description time [s]	13.029	0.422	0.383
Matching time [s]	5.90881	2.5402	2.7422

3 Conclusions

This paper proposed a novel two-step matching strategy and Center Symmetry FREAK (CS-FREAK) for feature description and matching. Compared with the previous proposed FREAK descriptor, CS-FREAK is quite different in the utilization of the neighborhood intensity information, encoding scheme, comparison rule and matching strategy. More specifically, it simplifies the 8-layer retina modal and more fully explores the local intensity relationships by considering the intensity order among the points in 4 directions around the sample points. Experimental results on various image transformations have shown that the proposed two-step matching strategy and CS-FREAK outperforms the state-of-the-art methods.

Acknowledgments. This work is supported by the National Science Foundation of China (61203189) and General Armament Department Pre-research Foundation (9140A01060411JB47-01).

References

1. Alahi, A., Ortiz, R., Vandergheynst, P.: FREAK: Fast retina keypoint. In: *Computer Vision and Pattern Recognition (CVPR)*, pp. 510–517. IEEE Press, New York (2012)
2. Lin, D., Xiliang, N., Tao, J., Shunshi, H.: Automatic registration of CCD images and IR images based on invariant feature. In: *Infrared and Laser Engineering*, vol. 40, pp. 350–354 (2011)
3. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. In: *International Journal of Computer Vision*, vol. 60, pp. 91–110. Kluwer Academic Publishers (2004)
4. Xingmiao, L., Shicheng, W., Jing, Z.: Infrared image moving object detection based on image block reconstruction. In: *Infrared and Laser Engineering*, vol. 1, pp. 176–180 (2011)
5. Liu, J., Zhang, T.: Recognition of the blurred image by complex moment invariants. In: *Pattern Recognition Letters*, vol. 26, pp. 1128–1138 (2005)
6. Ke, Y., Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors. In: *Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 506–513. IEEE Press, Washington, DC (2004)
7. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 1615–1630 (2005)
8. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006, Part I. LNCS*, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)

9. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: Brief: Binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010)
10. Leutenegger, S., Chli, M., Siegwart, R.Y.: BRISK: Binary robust invariant scalable keypoints. In: IEEE International Conference on Computer Vision (ICCV), pp. 2548–2555. IEEE Press, Barcelona (2011)
11. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: IEEE International Conference on Computer Vision (ICCV), pp. 2564–2571. IEEE Press, Barcelona (2011)
12. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part I. LNCS, vol. 2350, pp. 128–142. Springer, Heidelberg (2002)
13. Wang, Z., Fan, B., Wu, F.: Local intensity order pattern for feature description. In: IEEE International Conference on Computer Vision (ICCV), pp. 603–610. IEEE Press, Barcelona

Lung Fields Segmentation Algorithm in Chest Radiography

Guodong Zhang, Lin Cong, Liu Wang, and Wei Guo*

School of Computer, Shenyang Aerospace University
110136 Shenyang, China
{zhanggd, guowei}@sau.edu.cn, conglin_cl@sina.com,
aimengangela@sina.cn

Abstract. Accurate segmentation of lung fields in chest radiography is an essential part of computer-aided detection. We proposed a segmentation method by use of feature images, gray and shape cost, and modification method. The outline of lung fields in the training set was marked and aligned to create an initial outline. Then, dynamic program was employed to determine the optimal one in terms of the gray and shape cost in the six feature images. Finally, the lung outline was modified by Active Shape Model. The experimental results show that the average segmentation overlaps without and with feature images achieve 82.18% and 89.07%, respectively. After the modification of segmentation, the average overlap can reach 90.26%.

Keywords: Feature image, gray and shape cost, Active Shape Model.

1 Introduction

Lung cancer has become one of the most serious threatens to human life and health. Especially in recent years with environmental degradation caused by air pollution and a significant increase of smokers, the morbidity and mortality of lung cancer are growing fast. In American, an estimated 228,190 new cases of lung cancer and 159,840 deaths cases were expected in 2013, the death cases accounting for about 27% of all the cancer deaths[1]. In our country, the estimated 493,348 deaths in 2008[2], the mortality increased by 60% compared to 2004-2005. Chinese Cancer Prevention and Control Plan (2004-2010) considered lung cancer as an important issue [3]. Studies had shown that, if the cancer can be diagnosed and treated early, the number of the survival rate of patients with early stage lung cancer would be higher than 90% after 10 years [4]. Therefore, early detection and treatment of lung cancer becomes very critical.

With the development of computer-aided detection technology, the performance of lung cancer detection is greatly improved. Currently, chest radiography was used in most hospitals to detect lung cancer. The chest radiography has become the primary

* Corresponding author.

imaging modality in lung cancer detection because of its low cost and low dose radiation. In addition, the imaging equipment of chest radiography is simpler than other imaging modalities. Since accurate segmentation of lung fields is the basis of automatic detection of lung nodules, it has become one of the hotspots in the fields of medical image processing. The segmentation algorithm of lung fields based on the analysis of feature images was introduced by Xu et al. [5]. The top of lungs and the contour of chest cavity were determined by the second derivative of contour, and the right hemidiaphragm edges were determined by edge gradient analysis. The starting points were determined based on a “standard rule” to search for the left hemidiaphragm edges. Ginneken et al. [6] used active shape models, active appearance model and a multi-resolution pixel classification method for the segmentation of lung fields. Shi et al. [7] proposed a new deformable model by use of both population-based and patient-specific shape statistics to segment lung fields from serial chest radiographs. Soleymanpour et al. [8] used adaptive contrast equalization and non-linear filtering to enhance the original images. Then, an initial estimation of lung fields was obtained based on morphological operations, and it was improved by growing this region to obtain the accurate contour. Zhenghao Shi et al. [9] modified the conventional fuzzy c-means algorithm. Then the Gaussian kernel-based fuzzy clustering algorithm with spatial constraints was used for automatic segmentation of lung fields. Yan Liu et al. [10] proposed the lung segmentation algorithm based on the flexible morphology and clustering algorithm. In our study, we established an initial outline model and segmented the lung fields by use of gray and shape similarity information in feature images. However, the initial position of lung fields may be far away from the actual boundary of lung fields in some images. During segmentation with gray and shape similarity information, the lung outline cannot be covered by search area. Therefore, we modified the lung outline based on Active Shape Model (ASM) [11] algorithm to obtain a better segmentation result.

2 Establishment of an Initial Outline Model

The establishment of an initial outline model consists of three main sections: marking the training set, aligning the training set, and establishing an initial outline model. The process was described as follows.

2.1 Marking the Training Set

The points along the lung fields were used to mark the training set. The points included the following three classes: 1) the points with specific application; 2) the points without specific application, e.g. extreme points of the target in a specific direction or the highest point of the curvature; 3) the points between 1) and 2).

2.2 Aligning the Training Set

In order to compare the points in the same position from different training images, which need to be aligned to one another. The lung outlines of training images need to be aligned closely by scaling, rotation and translation operations.

2.3 Establishing an Outline Model

After aligning the shape of lung outlines in the training set, Principle Component Analysis was employed to determine statistical information of shape variations. Then an outline model was built, which can improve the efficiency of the algorithm.

The average shape of all the training images was assumed to $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$. The covariance matrix between the average shape and the training images after aligning was defined as $S = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T$. The value of N was the number of feature images. The eigenvalues and eigenvectors of the covariance matrix were calculated by $Sp_k = \lambda_k p_k$. Then the eigenvalues need to be sorted. The tors p_k correspond to the k largest eigenvalues λ_k .

The t principal eigenvectors need to be selected to form a new spindle system p_s . Then any shape belongs to the shape domain can be approximated by an average shape and weighted spindle system:

$$x = \bar{x} + p_s b_s \quad (1)$$

Where $p_s = (p_1 p_2 \cdots p_t)^T$ is a matrix composed with t tors, $b_s = (b_1 b_2 \cdots b_t)^T$ is the weight vector, b_s ensured the percentage of target object deformation determined by t eigenvalues accounting for the target object deformation determined by all eigenvalues is not less than V (V is general 0.98), and b_s should be limited to the condition $-3\sqrt{\lambda_k} \leq b_{s_k} \leq 3\sqrt{\lambda_k}$.

3 The Segmentation of Lung Fields Based on Gray and Shape Cost

In this paper, we used gray and shape cost simultaneously to make the gray and shape information of lung outline similar to the training images.

3.1 Feature Image

The variation of the image gray is prominent in feature image, therefore, the feature image was used to obtain the candidate points and the gray cost of each candidate point. Before obtaining the feature image of the original image, the Gaussian filter was required to reduce the effect of noise on segmentation results.

Six feature images were used in this paper: (1-2) x, y direction of the first order partial derivatives, which means x, y direction gray variation; (3-4) x, y direction of the second order partial derivatives, represents x, y direction of the gradation variation rate; (5) x, y direction of the mixed partial derivative; (6) x, y direction with the second order partial derivatives. The larger value indicates that the gray variation is fast, and the possibility of being the lung outline is large.

3.2 Similar Cost

Gray Cost. For each point of prior lung outline in test images, the degree of similarity between the gray vector of all pixels in search area of this point with six feature images and gray vector of the corresponding point in training images was computed, then the m points with maximum similarity were selected as the candidates of boundary. Degree of similarity of the point i can be evaluated by the cosine of the angle:

$$h_i = \sum_{j=1}^N \frac{(g_i^j)^T u_{g_i^j}}{\|g_i^j\| \times \|u_{g_i^j}\|} \quad (2)$$

Where g_i^j is the gray vector of the point i in feature image j , the gray vector can be described by the set of the gray of the points that located on a circle which centered at the point i with radius r_c , $u_{g_i^j}$ is the average of the gray vector of the corresponding point in feature image j , N is the number of feature images, which the value is 6. The larger h_i value of boundary candidate points in test images indicated that the higher similarity in gray distribution between this point and the corresponding point in training images.

Shape Cost. Shape cost of each boundary point was defined as the degree of similarity between shape feature of this point and average shape feature of the corresponding point in training images. The degree of similarity of the point i can be defined as:

$$f_i = \frac{(v_i)^T u_{v_i}}{\|v_i\| \times \|u_{v_i}\|} \quad (3)$$

Where v_i equals $p_{i+1} - p_i$ represents the shape feature of the point i , and p_i was the coordinate of boundary point i , u_{v_i} represents the average shape feature of the corresponding point in training images.

3.3 Segmentation by Use of Dynamic Programming

For the boundary point i in test images, once the m candidate points with smaller gray cost were selected, a $n \times m$ gray cost matrix M was constructed. Therefore, searching for the optimal outline thus translated in finding an optimal path through M . That means selecting an element from each row of the matrix M to constitute the optimal path [12]. In the process of finding the optimal path, the sum of gray and shape cost must be the largest:

$$J(k_1 \cdots k_n) = \sum_{i=1}^n h_i + \gamma \quad (4)$$

Where γ is the coefficient between gray and shape cost. To make the gray and shape cost played roughly the same role in searching the boundary points searching, the γ need to be adjusted properly.

4 Lung Outline Correction Based on Active Shape Model

In some images, the initial position may be too far away from the actual lung outline. When segmenting the lung outline by use of gray and shape similarity information, the lung outline may be not covered by search region. Therefore we need to adjust the lung outline, and improve the situation of some points getting stuck in a local optimum to obtain the better segmentation results.

In this phase, the lung outline of first segmentation was modified by the Active Shape Model. Firstly, we need to find the shape parameters to adjust the lung outline, and then the new lung outline was obtained, where the shape parameters can be obtained according to the ASM. Then, the degree of similarity of gray distribution in gradient direction between the new boundary points in all feature images from test sample and the boundary points from training sample must be considered. The larger degree of similarity indicated that the possibility being the optimal lung outline is great. These above two phases need to be carried out iteratively until obtain the most optimal lung outline.

5 Experimental Results and Analysis

5.1 Experimental Data

In our study, the experimental data used herein is the public database from the Japanese Society of Radiological Technology. It includes 247 Posterior-Anterior images, of which 93 images are normal, of which 154 images with nodules. The size of

the chest radiographs is 2048×2048 pixels, every pixel is 12 bits, and the scale of every pixel is $0.175 \times 0.175mm^2$.

5.2 Parameters Setting

When 1) the number of initial lung boundary candidate points are 30; 2) the coefficient between gray and shape cost is 10; 3) Right lung: the search region of the 4-16, 27-39 boundary points are the 60×60 square region centered at the point, the remaining boundary points are 100×100 square region; Left lung: the search region of 4-16, 36-39 boundary points are the 60×60 square region centered at the point, the remaining boundary points are 80×80 square region; 4) six feature images are used to search the lung outline. After modifying the lung outline, the optimal segmentation performance would be obtained and the average overlap of lung segmentation can reach 90.26%. The average overlap is a standard for measuring similarity between the final segmentation results and the initial marking outline.

5.3 Experimental Result

Segmentation Results with Different Searching Regions. In test images, the candidate points with larger gray cost of each boundary points need to be determined before searching the optimal lung outline, which were from the search region corresponding to each boundary point. Generally speaking, the gray distribution of different positions in chest radiographs are different, therefore, we need to classify the search region according to the different situations. The R_i represent the i boundary point on right lung and the L_i represent the i boundary point on left lung. There are 42 points both on left and right lung. Under the same conditions, the segmentation results with different search regions were given in Table I .

Table 1. Segmentation Results with Different Search Regions

R1-R3,R17-R26,R40-R42	L1-L3,L17-L35,L40-L42	Others	Overlap (%)
80	120	80	88.56
90	100	70	89.19
100	80	60	90.26

Segmentation Results under Different Conditions. The algorithm proposed in this paper would generate different segmentation results under different conditions. In this

Table 2. Segmentation Result under Different Conditions

Condition	Without feature images	With feature images	With feature images and the lung outline correction
<i>Overlap (%)</i>	82.18	89.07	90.26

section, we will show the segmentation performance under three conditions: without feature images, with feature images, with feature images and lung outline correction. The segmentation results were given in Table II. Fig.1 shows the segmentation performance of lung outline under the three conditions. As can be seen from the Fig.1, the lung segmentation algorithm based on feature images and lung outline correction proposed in this paper can obtain the optimal segmentation results.

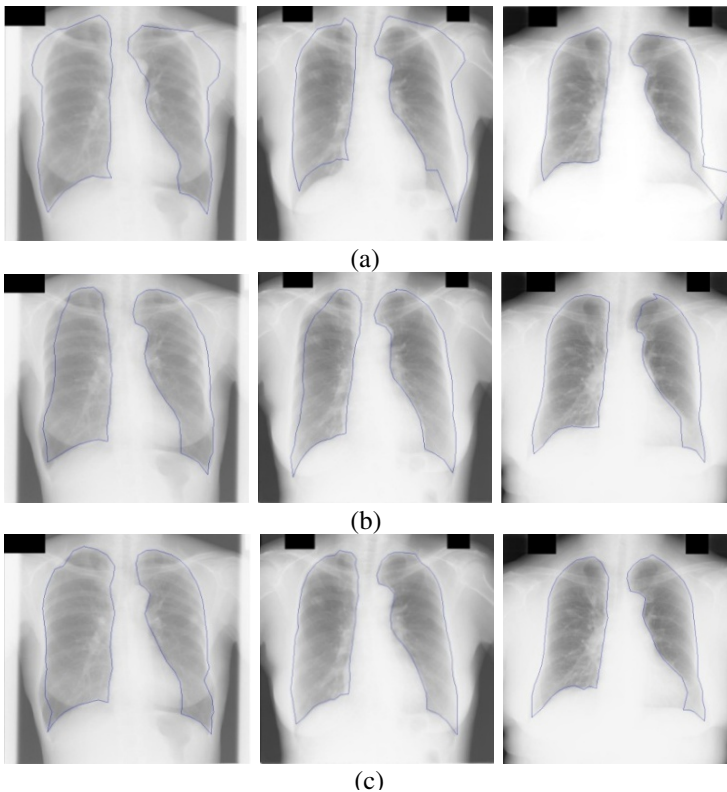


Fig. 1. The Segmentation results without feature images and with feature images are shown in (a) and (b), respectively. The final segmentation result with feature images and lung outline correction is shown in (c).

6 Conclusion

In this paper, we proposed a segmentation algorithm based on feature images, gray and shape cost, which can improve the performance of the segmentation of lung outline. Moreover, modification of the lung outline based on ASM was employed to overcome the limits of the search regions not covered the real lung fields, and further improved the performance of the segmentation of lung fields.

Acknowledgements. This work was supported in part by National Natural Science Foundation of China (No.61373088), the PhD Start-up Fund of National Science Foundation of Liaoning Province (No.20131086), National Aerospace Science Foundation (No.2013ZE54025), Shenyang Science and Technology Foundation (No.F13-316-1-35), and the PhD Start-up Fund of SAU (No.13YB16).

References

1. Cancer Facts and Figures 2013. The American Cancer Society, Atlanta (2013)
2. Li, Y., Dai, M., Chen, L., Zhang, S., Chen, W., Dai, Z., et al.: Study on the Estimation of Lung Cancer Mortality of Provincial Level Chinese. *China Oncology*, 120–126 (2011)
3. Lei, T.: The Ten Most Common Cancer Mortality and Composition in China. *China Cancer* 12, 801–802 (2010)
4. The International Early Lung Cancer Action Program Investigators Survival of Patients with Stage I Lung Cancer Detected on CT Screening. *The New England Journal of Medical* 355, 1763–1771 (2006)
5. Xu, X.X., Dio, K.: Image Feature Analysis for Computer-aided Diagnosis: Detection of Right and Left Hemidiaphragm Edges and Delineation of Lung Field in Chest Radiographs. *Medical Physics* 23, 1613–1624 (1996)
6. Ginneken, B., Stegmann, M., Loog, M.: Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *Med. Image Anal.* 10, 19–40 (2006)
7. Shi, Y., Qi, F., Xue, Z., Chen, L., Ito, K., Matsuo, H., Shen, D.: Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics. *IEEE Transactions on Medical Imaging* 27, 481–494 (2008)
8. Soleymanpour, E., Pourreza, H.R., Ansaripour, E., Yazdi, M.: Fully automatic lung segmentation and rib suppression methods to improve nodule detection in chest radiographs. *J. Med. Signals Sens.* 1, 191–199 (2011)
9. Shi, Z., Zhou, P., He, L., Nakamura, T., Yao, Q., Itoh, H.: Lung segmentation in chest radiographs by means of Gaussian kernel-based FCM with spatial constraints. In: ICNC-FSKS, Tianjin, China (2009)
10. Liu, Y., Qiu, T., Guo, D.: The lung segmentation in chest radiographs based on the flexible morphology and clustering algorithm. *Chinese Journal of Biomedical Engineering* 26, 684–689 (2007)
11. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models — their training and applications. *Computer Vision Image Understand* 61, 38–59 (1995)
12. Zhang, Y., Chen, X., Hao, X., Xia, S.: Dynamic programming algorithm for pulmonary module CTimages based on counter supervision. *Chinese Journal of Scientific Instrument* 33, 25–27 (2012)

Automatic Classification of Human Embryo Microscope Images Based on LBP Feature

Liang Xu¹, Xuefeng Wei^{1,2}, Yabo Yin³, Weizhou Wang⁴,
Yun Tian^{3,*}, and Mingquan Zhou³

¹ College of Information Engineering, Huanghuai University, Zhumadian, China

² College of Information Engineering, Wuhan University of Technology, Wuhan, China
xl66315@163.com, jkxwlsy@126.com

³ College of Information Science & Technology, Beijing Normal University, Beijing, China
yinyabo0612@163.com, tianyun@bnu.edu.cn, mqzhou@bnu.edu.cn

⁴ Assisted reproductive medical center, Navy General Hospital, PLA, Beijing, China
wangweizhou12@126.com

Abstract. It is significant in-vitro fertilization (IVF) to automatically evaluate the implantation potential for embryos with a computer. In this essay, an automatic classification algorithm based on local binary pattern (LBP) feature and the support vector machine (SVM) algorithm is presented to classify the embryo images which will suggest whether the image is suitable for the implantation. The LBP operator is first time to be used to extract the texture feature of embryo images, and it is verified that the feature has the capacity of making two types of images linearly separable. Furthermore, a classifier based on the SVM algorithm is designed to determine the best projection direction for classify embryo images in the LBP feature space. Experiments were made with 6-fold cross validation over 185 images, and the result demonstrates that the proposed algorithm is capable of automatically classifying the embryo images with accuracy and efficiency.

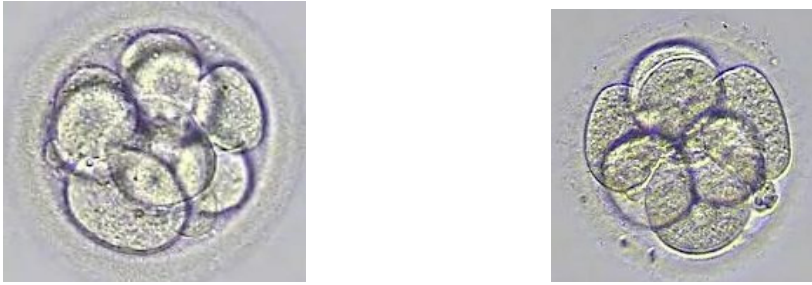
Keywords: Embryo image, classification, local binary pattern and support vector machine.

1 Introduction

With the development of science and technology, IVF (in-vitro fertilization) technology has become one of the main methods of treating infertility[1]. Although IVF technology has made great progress, the efficiency of IVF still needs improvement. A great challenge that the embryologists face is how to recognize the most viable embryo to be implanted. Currently, the embryologists evaluate the implantation potential of embryos by visual examination and their evaluation is totally subjective. To solve this problem, automatic evaluation of embryo's quality with the help of embryo images before transferring by computer-aided method and then selecting the optimal

* Corresponding author.

embryo to transfer is currently a hot research field for human reproductive [2-4]. The embryo morphology during the third day or the fifth day after fertilization can be used to evaluate the implantation potential[5-7]. Fig. 1 shows the morphology of an embryo during the third day. Fig. 1(a) shows an embryo that successfully becomes a baby, and Fig. 1(b) shows an embryo that fails to develop into a baby. Usually, the texture of successful pregnancy embryos is relatively regular for embryo images, and the texture of the embryos that fail to be implanted is rather disorganized.



(a) an embryo leading to successful pregnancy (b) embryo that fails to give birth

Fig. 1. Embryo images in the third day after fertilization

In this essay, we mainly focus on the embryo viability by exploiting their morphological features during the third day after fertilization, which can be considered as the classification of embryos, and feature extraction of embryo images and classifier design are involved[8]. Özkaya[9] studied the impact of maternal age, the number of eggs, the size of blastomeres and thickness of zona pellucida on embryo's classification. Scott et al[10] used the nuclear morphology of embryos to do classification. Hnida et al[11] made use of multi-level morphological analysis method to classify embryos. However, these methods are difficult to extract and quantify feature accurately. Due to the abundant texture of embryo images, it is expected to be more reasonable to quantify the characteristics of embryo images. As a texture descriptor, the LBP descriptor is effective for delineating the local texture feature of images. The descriptor compares each point with its neighboring pixels, and the comparing result is saved as a binary number. Due to its powerful discrimination and simple calculation, the LBP operator has been applied in fingerprint recognition, face recognition, license plate recognition and other areas[12,13]. The most important advantage of the LBP is its robustness to the changes of image intensities; this advantage makes the LBP more suitable for embryo images taken by HMC (huffman modulation contrast). Moreover, thanks to the simple calculation, this operator can be used for real-time analysis. Therefore, we attempt to employ this operator to do feature extraction on embryo images.

In addition, how to design the classifier is another key issue for the embryo image evaluation[14]. The SVM (support vector machine) is a supervised learning method, and it tries to establish a hyper-plane with the maximum interval in the feature

space via learning and training the labeled samples, then the unlabelled data can be classified by the plane. The SVM is widely used for statistical classification, and it can solve the learning problem of samples with small number[15]. Due to the limited number of embryo image samples, we will develop a classifier base on the SVM algorithm to categorize the embryo images.

2 Classification Algorithm

The SVM algorithm[5] is capable of solving the classification problem of samples with small number, nonlinear and high dimensional pattern recognition, and an important property is that it can be extended to the case of the non-linear inseparable situation. Based on the algorithm, the optimal classification plane can be defined as in the following formula: let the linearly separable sample set as $(x_i, y_i), i = 1, 2, \dots, n$, $x_i \in R^d$, x_i represents a mode of d-dimensional space; n denotes the number of modes; $y_i \in \{+1, -1\}$, y_i denotes the class label of the sample. In d -dimensional space, the general form of the linear discriminant function is $g(x) = \vec{w} \cdot \vec{x} + b$, and the classification hyper-plane equation can be expressed as

$$\vec{w} \cdot \vec{x} + b = 0 \quad (1)$$

For the equation $g(x) = \vec{w} \cdot \vec{x} + b$, it should make all samples of the two types meet $|g(x)| \geq 1$, i.e., the nearest samples from the classification plane should meet $|g(x)| = 1$. Assume that the two hyper-planes L_1 and L_2 respectively pass the points from two sample sets, and they are nearest from the classification plane H and parallel to it. The distance between L_1 and L_2 is defined as the classification interval $d_{interval}$ of the two classes. As a result, the optimal classification plane H not only should separate the two kinds of samples without error, but also needs to make the classification interval maximum.

According to the classification surface equation, the classification interval $d_{interval}$ can be calculated as $2/\|\mathbf{W}\|$. Making $d_{interval}$ maximum means $\|\mathbf{W}\|^2$ to be minimized. Also, if the classification plane H can correctly classify all samples, it needs to meet the following condition:

$$y_i[(\vec{w} \cdot \vec{x}) + b] - 1 \geq 0 \quad (2)$$

Thus, the problem of seeking the optimal classification H plane can be converted to the problem of minimizing the following function with the constraint $y_i[(\vec{w} \cdot \vec{x}) + b] - 1 \geq 0$:

$$\min \quad \varphi(\vec{w}) = \frac{1}{2} \|\vec{w}\|^2 = \frac{1}{2} \vec{w} \cdot \vec{w} \quad (3)$$

The above problem can be solved by defining its Lagrangian function as follows:

$$L(\vec{w}, b, \vec{\alpha}) = \frac{1}{2}(\vec{w} \cdot \vec{w}) - \sum_{i=1}^n \alpha_i \{ y_i [(\vec{w} \cdot \vec{x}_i) + b] - 1 \} \quad (4)$$

where α_i denotes the Lagrange coefficient, and $\alpha_i \geq 0$. The partial derivative about w and b can be obtained, and are set to zero:

$$\sum_{i=1}^n y_i \alpha_i = 0, \vec{w} - \sum_{i=1}^n y_i \alpha_i \vec{x}_i = \vec{0}, \quad \alpha_i \geq 0 \quad (5)$$

The Eq.(5) is substituted into the Eq.(4), and the original problem can be converted to maximize the following function with constraints $\sum_{i=1}^n y_i \alpha_i = 0$ and $\alpha_i \geq 0, i = 1, 2, \dots, n$:

$$\max \quad Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (\vec{x}_i \cdot \vec{x}_j) \quad (6)$$

Assume that α_i^* is the optimal solution to the problem, then w^* can be obtained as $w^* = \sum_{i=1}^n \alpha_i^* y_i \vec{x}_i$. In fact, Eq. (6) can be considered as a quadratic programming problem, and it has a unique solution. According to Kuhn-Tucker conditions, the solution of this optimization problem needs to satisfy the following equation:

$$\alpha_i (y_i (\vec{w} \cdot \vec{x}_i + b) - 1) = 0, i = 1, \dots, n \quad (7)$$

For the non-support vectors, $y_i (\vec{w} \cdot \vec{x}_i + b) - 1$ in Eq.(8) will be greater than 0, and only those α_i corresponding to the support vectors are 0. After determining the optimal solution w^* , any support vector is substituted into $y_i [(\vec{w} \cdot \vec{x}_i) + b] - 1 = 0$ and b can be obtained. To avoid error, the parameter b is usually set as $b^* = -\frac{1}{2} \sum_{SV_s} y_i \alpha_i^* (s_1^T x_i + s_2^T x_i)$, where s_1, s_2 denote support vectors, and SV_s

denotes the support vector sets. After parameters w^* and b^* are determined, the following optimal classification function can be obtained:

$$\begin{aligned} f(\vec{x}) &= \text{sgn}\{(\vec{w}^* \cdot \vec{x}) + b^*\} \\ &= \text{sgn}\left\{\sum_{i=1}^n \alpha_i^* y_i (\vec{x}_i \cdot \vec{x}) + b^*\right\} \end{aligned} \quad (8)$$

Since the α_i value corresponding to the non-support vectors are 0, the summation computation in Eq. (8) can be greatly reduced.

3 Experiment and Analysis

3.1 Feature Selection

We extracted two kinds of features on 185 embryos images from Assisted Reproductive Center of Navy General Hospital, PLA (the number of samples that are successfully pregnant is 47, the number without success to conceive is 138) has been used to make classification experiments. The first kind of feature is the five order central moments of embryo images, namely using each first five central moments in horizontal and vertical direction to classify embryos. The central moments contain texture - related information of the change intensity of images, and they can be used to classify the embryo images[3]. The corresponding central moment vector of each sample is 12-dimensional. The second kind of feature is the LBP feature. The LBP operator can be used to measure and extract local texture information of images, and it is illumination invariant. In this experiment, the LBP vector of each sample is 256-dimensional.

To verify the classification algorithm, we used six cross-over trials to design the classifier. Data were randomly divided into six groups. One group was chosen as the test data each time, and the remaining five groups were employed to train the classifier.

3.2 Classifier Design and Analysis

3.2.1 Central Moment

When all five central moments of the original images are used to do classification, the support vector indexes are 2,5,7,171, respectively. Based on the projection direction w that is $[0.0009, 0.0001, 0.0007, 0.0136, 0.0619, -0.0043, 0.0009, 0.0002, 0.0018, 0.0379, 0.1684, -0.0036] \times 10^{-3}$, the b^* values corresponding to the four support vectors are all -2.5384 . So the final decision function took the offset b as -2.5384 . We used this classifier to classify the embryo images and the average accuracy rate was 62.16%. Fig. 2 shows the projection result of samples to one-dimensional space

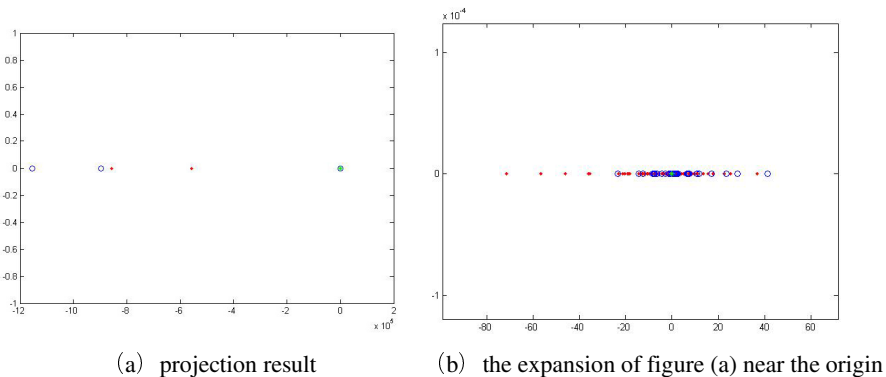


Fig. 2. The adding result of samples' projection onto the classification plane's normal vector and offset for central moments

Table 1. The cross validation result using five central moments

	accuracy (%)						mean(%)	variance
	group1	group2	group3	group4	group5	group6		
1	77.42	45.16	80.65	54.84	74.19	66.67	66.49	0.0162
2	70.97	74.19	70.97	64.52	51.61	60.00	65.38	0.0060
3	58.06	70.97	51.61	—	64.52	70.00	63.03	0.0054
4	—	48.39	58.06	80.65	58.06	60.00	61.03	0.0113
5	54.84	80.65	61.29	58.06	74.19	76.67	67.62	0.0098

(for better display, we added a vertical axis), the horizontal axis indicates the result of samples' projection onto the classification normal vector adding offset, and the zero point in the horizontal axis represents the threshold of classification. Fig. 2(b) is the expansion of Fig. 2(a) near the origin. As one can see from the figure, the samples are mixed up after the projection and cannot be classified by a reliable plane which illustrates that the first five order central moments of embryo images are linearly inseparable.

The cross validation result using five central moments as features is shown in Table 1, and the experiment was repeated five times. In the Table, "—" represents failure to properly find a sufficient number of support vectors. The mean column shows that our classification performance is general on the first five central moments, but the variance column indicates that the classification algorithm is relatively stable.

3.2.2 LBP Operator

When all LBP values are used to classify embryo images, the number of support vector indexes is 114. The offset b is -8.0353 . The decision function was substituted back to the original data to do the classification; and the classification accuracy rate was 100% and the number of misclassified samples was 0. This indicates that the original data is separable. Fig.3 shows the projection result of samples to one-dimensional space (for better display, we added a vertical axis). As one can see from the figure, sample points after the projection are of very strong separability which shows that the use of our classifier to classify LBP data is feasible. The result of Fig. 4 indicates that the LBP features of the embryo images are linearly separable, and the case of failure classification does not exist. Table 2 shows the cross validation classification accuracy through the LBP feature. It is evident that the accuracy has a slight increase with respect to the five order central moments, which seems to be not cohere with the fact that LBP data is linearly separable. The main reason is that each calculated normal vector and offset volume of the classification equation are quite different after the sample data randomization. This suggests that the selection of training samples has a great impact on the classification plane which is determined by the quite small number of training samples.

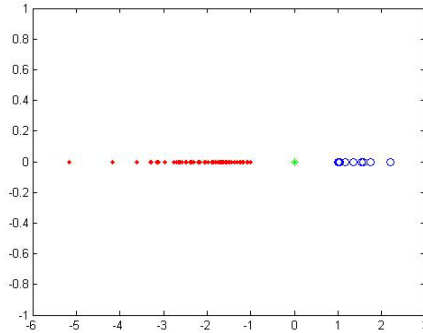


Fig. 3. The adding result of samples' projection onto the classification plane's normal vector and offset for LBP

Table 2. The cross validation result using the LBP feature

	accuracy (%)						mean(%)	variance
	group1	group2	group3	group4	group5	group6		
1	80.65	61.29	54.84	64.52	67.74	63.33	65.39	0.0062
2	35.48	70.97	64.52	67.74	80.65	73.33	65.45	0.0205
3	64.52	64.52	67.74	58.06	58.06	63.33	62.07	0.0013
4	64.52	64.52	54.84	61.29	70.97	80.00	66.02	0.0062
5	74.19	64.52	74.19	74.19	67.74	66.67	70.25	0.0016

4 Conclusion

In this research paper we present a LBP-based automatic classification algorithm on microscopic images of human embryos. It verified the validity of LBP as the feature to do classification representing the local texture of embryo images, and it can be make the two types of embryo images linearly separable; At the same time, when combined with the SVM algorithm, the optimal projection direction for effectively classifying LBP features was determined. The proposed algorithm is important for computer-aided embryo transfer. In our further study, more training samples will be collected for training purposes to aquire the best performance of the classifier.

References

1. Santos Filho, E., Noble, J.A., Wells, D.: A Review on Automatic Analysis of Human Embryo Microscope Images. *The Open Biomedical Engineering Journal* 4, 170–177 (2010)
2. Siristatidis, C., Pouliakis, A., Chrelias, C., Kassanos, D.: Artificial Intelligence in IVF: A Need. *Systems Biology in Reproductive Medicine* 57, 179–185 (2011)

3. Paternot, G., Debrock, S., De Neubourg, D., D'Hooghe, T.M., Spiessens, C.: Semi-automated Morphometric Analysis of Human Embryos Can Reveal Correlations between Total Embryo Volume and Clinical Pregnancy. *Human Reproduction* 28(3), 627–633 (2013)
4. Guh, R., Wu, T.J., Weng, S.: Integrating Genetic Algorithm and Decision Tree Learning for Assistance in Predicting in Vitro Fertilization Outcomes. *Expert Systems with Applications* 38(4), 4437–4449 (2011)
5. Bendus, A.E.B., Mayer, J.F., Shipley, S.K., Catherino, W.H.: Interobserver and Intraobserver Variation in Day 3 Embryo Grading. *Fertility and Sterility* 86(6), 1608–1615 (2006)
6. Desai, N.N., Goldstein, J., Rowland, D.Y., Goldfarb, J.M.: Morphological Evaluation of Human Embryos and Derivation of an Embryo Quality Scoring System Specific for Day 3 Embryos: A Preliminary Study. *Human Reproduction* 15(10), 2190–2196 (2000)
7. Santos Filho, E., Noble, J.A., Poli, M., Griffiths, T., Emerson, G., Wells, D.: A Method for Semi-automatic Grading of Human Blastocyst Microscope Images. *Human Reproduction* 27(9), 2641–2648 (2012)
8. Morales, D.A., Bengoetxea, E., Larrañaga, P., García, M., Franco, Y., Fresnada, M., Merino, M.: Bayesian Classification for the Selection of in Vitro Human Embryos Using Morphological and Clinical Data. *Computer Methods and Programs in Biomedicine* 90(2), 104–116 (2008)
9. Özkaya, A.U.: Assessing and Enhancing Machine Learning Methods in IVF Process: Predictive Modeling of Implantation and Blastocyst Development, PH.D thesis, Boğaziçi University (2011)
10. Scott, L.A., Smith, S.: The Successful Use of Pronuclear Embryo Transfers the Day Following Oocyte Retrieval. *Human Reproduction* 13(4), 1003–1013 (1998)
11. Hnida, C., Engenheiro, E., Ziehe, S.: Computer-controlled, Multilevel, Morphometric Analysis of Blastomere Size as Biomarker of Fragmentation and Multinuclearity in Human Embryos. *Human Reproduction* 19(2), 288–293 (2004)
12. Ojala, T., Pietikäinen, M., Harwood, D.: Performance Evaluation of Texture Measures with Classification Based on Kullback Discrimination of Distributions. In: *Proceedings of the 12th IAPR International Conference on Pattern Recognition, ICPR 1994, vol. 1*, pp. 582–585 (1994)
13. Ojala, T., Pietikäinen, M., Harwood, D.: A Comparative Study of Texture Measures with Classification Based on Feature Distributions. *Pattern Recognition* 29, 51–59 (1996)
14. Uyar, A., Bener, A., Ciray, H., Bahceci, M.: A Frequency Based Encoding Technique for Transformation of Categorical Variables in Mixed IVF Dataset. In: *Proceedings Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2009*, pp. 6214–6217 (2009)
15. Erişti, H., Demir, Y.: A New Algorithm for Automatic Classification of Power Quality Events Based on Wavelet Transform and SVM. *Expert Systems with Applications* 37(6), 4094–4102 (2010)

Robust and Accurate Calibration Point Extraction with Multi-scale Chess-Board Feature Detector

Yang Liu¹, Yisong Chen², and Guoping Wang^{2,*}

¹Shenzhen Graduate School, Peking University

²Graphics Lab of EECS, Peking University

Beijing Engineering Technology Research Center of Virtual Simulation and Visualization
{lypku, chenys, wgp}@pku.edu.cn

Abstract . Chess-board grid has been widely used for camera calibration and the associated feature point extraction algorithm draws much attention. In this paper, a multi-scale chess-board feature point detector is proposed, along with a chess-board matching algorithm for a specific marker used in our 3D reconstruction system. Experiments show that our method is more robust and accurate compared to commonly used approaches.

Keywords: Multi-scale chess-board corner detection, camera calibration, 3D reconstruction, calibration point extraction.

1 Introduction

Modern image-based 3D reconstruction systems reconstruct 3D models from a series of pictures capturing the objects. In general a specific chess-board grid is printed beforehand and put into the scene to help calibrate the cameras.

The chess-board grid we use for camera calibration consists of two parts (As shown in **Fig. 1**): a blank square area in the middle for placing objects to be reconstructed, and a chess-board grid region around the blank area. In Fig. 1b the feature points are circled in green. The four corners of the grid are designed with distinctive style for identifying global orientation.

To ensure the integrity and accuracy of reconstruction, we have to capture the objects in as many views as possible and make sure they are well-focused. In most circumstances, the chess-board is arbitrarily placed, as shown in **Fig. 2**, and it often occurs that the board is ill-focused. These make it difficult to extract the chess-board vertices robustly and precisely. Besides, the variation of image size, the location of the board and the features of the objects increase the difficulty of calibration feature extraction as well.

In this paper, we propose a multi-scale chess-board feature detector for robust and precise detection of the calibration markers to meet these challenges. First, we use this

* Corresponding author.

detector to obtain all chess-board features in the image. Then we use the gradient information of the image to help eliminate outliers and organize the true vertices with a brand new bidirectional growth algorithm. Finally we map the features to 3D locations in the board to fulfill the camera calibration. As it shows in the experiments, our detector is quite stable and precise, even in extreme conditions like imperfect focus and ill-illuminations. Our method is completely automatic and performs better than commonly used methods.

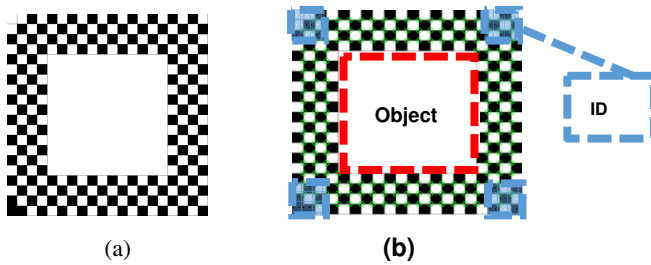


Fig. 1. Chess-board marker designed in this paper. (a) calibration chess-board (b) different parts consisted in this calibration board.



Fig. 2. Chess-board plane and the image plan have a large angle

2 Related Work

There are various published techniques for finding the chess-board vertices. H. Moravec [1] designed an algorithm which locates a corner by computing the gradients along eight directions with the anticipation that large response can be found along edges, but it is very prone to noises. Harris corner [2] is proposed based on [1] with an additional non-maximum suppression. F. Mokhtarian and R. Suomela [3] detected corners through curvature scale space with the help of Canny [4]. They didn't make use of the properties of the feature and are likely to suffer information loss by wide kernel filtering.

E. Rosten and T. Drummond [5] designed a quick corner detecting method called FAST which takes samples around a pixel around and determine whether it is a corner.

The Harris and Stephens [2] corner detector is adopted in [6] to locate a grid, before Hough transform is employed to constraint linearity and discard false responses.

This scheme performed badly when the grid is distorted because of optical distortion or on a non-planar surface.

Yu and Peng[7] adopt a pattern-match method to find specific features by measuring the correlation over all the image. This method fails when the grid is rotated relative to the patterns in store.

Sun et al. [8] describe a method which places a rectangular or circular window over every position of the image before achieving a 1D binarized vector along the perimeter. The positions where the vector has four regions are determined to be chess-board vertices. This method is somehow rather slow and prone to noises. The performance also relies on the result of binarization.

Shen Cai and Zhuping Zang [9] designed a deformed chessboard pattern for automatic camera calibration, but the precision and robustness of their method needs to be improved.

Stuart Bennett and Joan Lasenby [10] offer an instructive approach where they emphasize the importance of chess-board detector and design ChESS which is both simple and quick. It is rather accurate but tends to produce much more false features. The single-scale property also aggravates its limitation.

Besides, some open source labs such as OpenCV [11] and Matlab [12] tools are widely used for chess-board grid detection for their convenience, but they both have some problems concerning to accuracy and robustness.

3 Corner Detection Algorithm

Our chess-board marker detection algorithm for camera calibration mainly consists of two steps. First, we obtain all the chess-board features with the help of the multi-scale chess-board feature detector and eliminate outliers from them. Then we organize the vertices found in former step and map them to 3D locations in the marker. The detailed algorithm will be introduced by these two steps.

3.1 Detecting Chess-Board Vertices

In this section, we propose and justify several properties of the chess-board **calibration pattern**, based on which we design a single-scale annular chess-board **feature detector** and then develop it into a multi-scale detector to ensure the robustness and precision of the detection.

3.1.1 Annular Chess-Board Detector

When we put a sample circular window over a chess-board vertex, the points on opposite sides **of a diameter** tend to have similar intensity and those on perpendicular **radii** should have very different intensity (**Fig. 3**). Based on this observation, we build an annular chess-board detector which has two kinds of energies: E_{sym} and E_{dif} .

Unlike [10] where the two properties are combined into one, we find it produces much fewer false features if considering them separately.

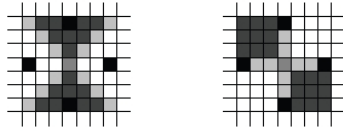


Fig. 3. points on opposite and perpendicular sides [10]

As illustrated in **Fig. 4**, taking c_i as a sample point on the circular window, c'_i is the point on the opposite side and b on the perpendicular side, N is the number of sampling points on the semicircle, these two energies can be formulated as:

$$E_{sym} = \sum_{i=0}^N (c_i - c'_i)^2 \tag{1}$$

$$E_{dif} = \sum_{i=0}^N (c_i - b)^2 \tag{2}$$

The former energy should be small over **the semicircle** while the latter relatively larger.

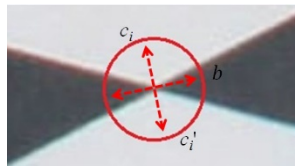


Fig. 4. Chess-board vertex

3.1.2 Multi-scale Annular Chess-Board Detector

If we adopt single-scale detector over the image, the radius of sampling window must be carefully chosen. Small rings over the blurred region of a vertex can't provide discriminating features while too large rings risk sampling pixels from unrelated squares which doesn't form the current features (**Fig. 5**). This creates a dilemma when processing different images. It is both unrealistic to apply only one sampling radius over all images and also inconvenient to choose an appropriate radius for every image.



Fig. 5. sampling circle (a) ideal vertex (b) vertex with blurriness(c)sampling circles of different sizes

To solve these problems, we employ a pragmatic multi-scale scheme which carries out sampling windows of different radius over a position, just like [8] (Fig. 6). The construction of multi-scale annular chess-board detector is quite quick using similar method to SURF [13]; however, it's much slower than that of the single-scale detector. We propose some strategies to speed up the detection in 3.1.4, which makes our detector more competitive.

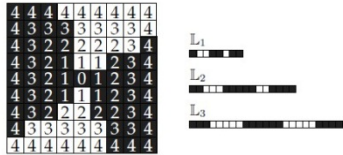


Fig. 6. Sun's layers [8]

3.1.3 Feature Selection

Our detector is quite distinctive and can almost detect all the visible chess-board vertices, however, false features might exist. Eliminating the false features is crucially important. In our study, the vertices must satisfy these constraints while false ones can't satisfy all of them:

1. $E_{sym} < E_1$, E_1 is an upper bound we set
2. $E_{dif} > E_2$, E_2 is a lower bound we set
3. There must be four color regions on the circular window
4. The path between two neighboring vertices must be along the edge of a chess-board square

The construction of multi-scale detector over every pixel is time consuming, so we start with a small radius and check the first three constraints, and increment it gradually if it fails the constraint 2 and 3 while satisfies constraint 1. Constraint 4 is employed by computing the sobel response of the image (Fig. 7). Large response tends to be along the edges of the square. Non-minimum suppression of E_{sym} is used over a small region where several features exist.

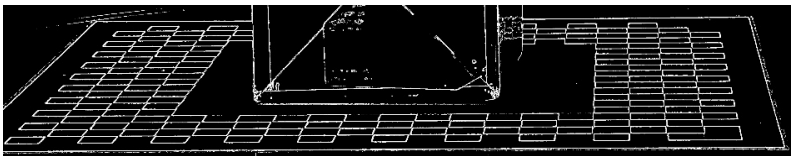


Fig. 7. Sobel response of the image

As shown in Fig. 8, sufficiently many vertices are detected while few false ones left.



Fig. 8. Vertices detected

3.1.4 Speed-Up Strategy

We propose two strategies to speed up the detection scheme. Instead of constructing our detector over every pixel of the image, we consider those positions where sobel responses are relatively high. Anyway, the threshold should be low enough to not risk eliminating the accurate positions of the vertices when blurring exists. Down-sampling the image can notably accelerate the computation, but an additional step to find the accurate positions of the vertices in the original image should be added.

3.2 Mapping the Vertices to 3D Board

After the vertices are detected, we must find out their locations in the 3D board before camera calibration. As shown in **Fig. 9**, the vertices form 12 lines along the periphery of the marker and each 3 lines make up a group. Usually at most two groups are invisible because of the occlusion of the objects, so we have at least two intersecting groups to locate an ID in the corner of the board. Hough transform [14] is the most widely used method when dealing with line fitting problems, but it is at great disadvantage when there's remarkable camera distortion. Inspired by [6], we design a bidirectional growth algorithm to fit these lines.

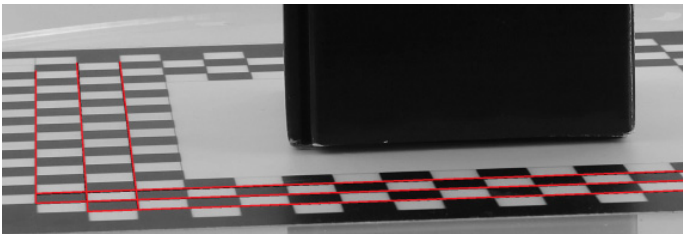


Fig. 9. 12 lines along the periphery

The algorithm is demonstrated in **Fig. 10**. For a vertex A, we find the closest vertex, named B, with which can form a path along the edge of the square (consider the sobel response). Put A and B in the inlier set and make growth along \overline{AB} : If another vertex and the last two inlier vertices lie along the edges of square, then put it into the set and keep looking for the next one . This process ends before the growth along the other direction \overline{BA} starts. One-direction growth is enough when the vertices are ordered along the x or y axis.

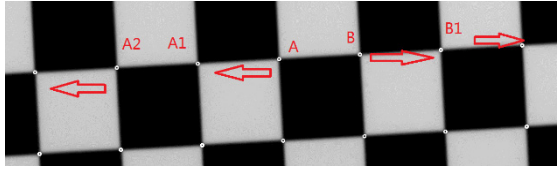


Fig. 10. Bidirectional growth algorithm

The lines successfully fit are illustrated in **Fig. 11.**



Fig. 11. Results of line fitting

The four ID codes in the corners of the grid are different from each other and have quite stable and distinctive appearances that are easy to be recognized. The 3D coordinates of the vertices detected are then obtained by using the method in [15].

4 Experimental Results

To show the contributions of our approach, we have performed the following two experiments: First, to evaluate the precision we compare the average reprojection error of the chess-board vertices detected by our detector with that by Harris and ChEss; Second, to evaluate the robustness we compare the chess-board recognition success rate of our method with that of J. Sun [16] and de la Escalera. We conduct an additional experiment over the images with different illumination, image sizes and camera poses to prove the validity of our method.

4.1 Precision

The dataset we use for this experiment consists of 20 groups (each group has about 20 images captured around the object on the chess-board) of images with good illumination.

These images can also be divided into 3 categories according to the angle between the image plane and the chess-board plane: high position (the angle is less than 30 degree), middle position (the angle is less than 60 degree and more than 30 degree), low position (the angle is more than 60 degree). As **stated** before, the camera focuses mainly on the objects to be reconstructed, so the **calibration pattern** is often captured with imperfect focus, which often cause blurriness in the condition of low position. We separately measure the average distance between the projected points of the 3D vertices and the vertices detected with Harris, ChEss and our detector. As shown in **Fig. 13**, our detector performs as well as ChEss, while much better than Harris, the average reprojection error is less than 1 pixel in all three conditions. We can also see that imperfect focus affects the accuracy, but our detector performs much more stable than Harris. ChEss is rather accurate but tends to produces much more false features and the sing-scale property also aggravates its limitation.

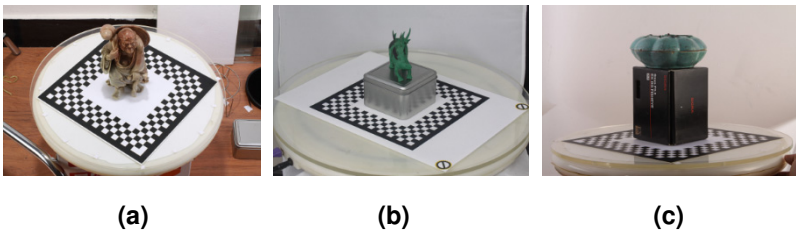


Fig. 12. Three categories divided by camera positions (a) high position (b) middle position (c) low position

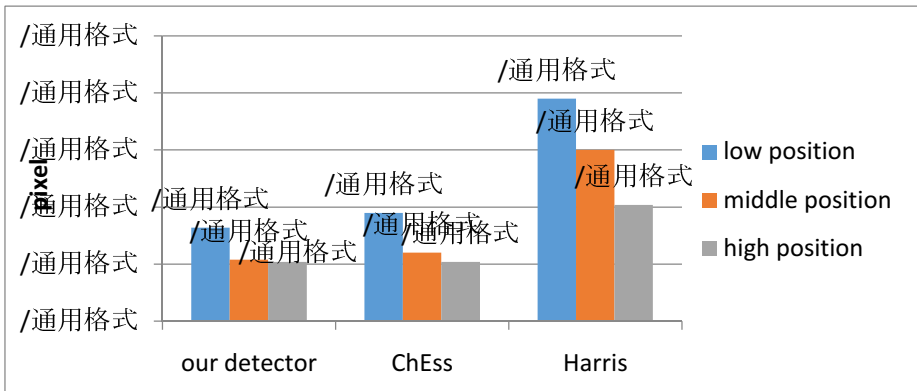


Fig. 13. Reprojection error of ChEss, Harris and our detector

Successful chess-board detection is to extract enough vertices, while discarding the false ones, and map them to 3D locations in the board to complete the camera calibration. There are generally three kinds of approaches in this field: J. Sun’s method using LSC and line fitting scheme, Escalera method combining Harris with Hough transformation.

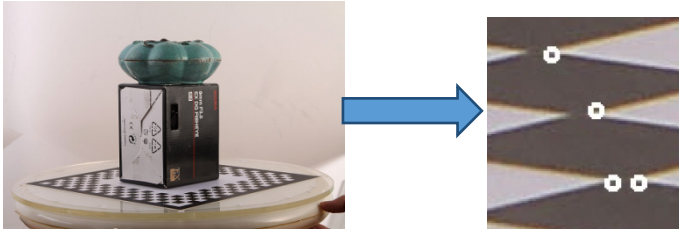


Fig. 14. An example where Harris corner deviants from true vertices

Our method differs with these two approaches. The comparison will be conducted in this experiment.

The dataset we use here consists of 40 groups of images which can also be divided into three categories like 4.1. We measure the chess-board recognition success rate and show some examples where these three methods would fail.

Fig. 15 shows that our method succeeds to calibrate the most images, which means it's the most robust among the four methods. The other three perform rather badly in the condition of low camera position. The line fitting scheme of de la Escalera is at a disadvantage when distortion or blurring exists. J. Sun's method has difficulty with discarding false vertices.

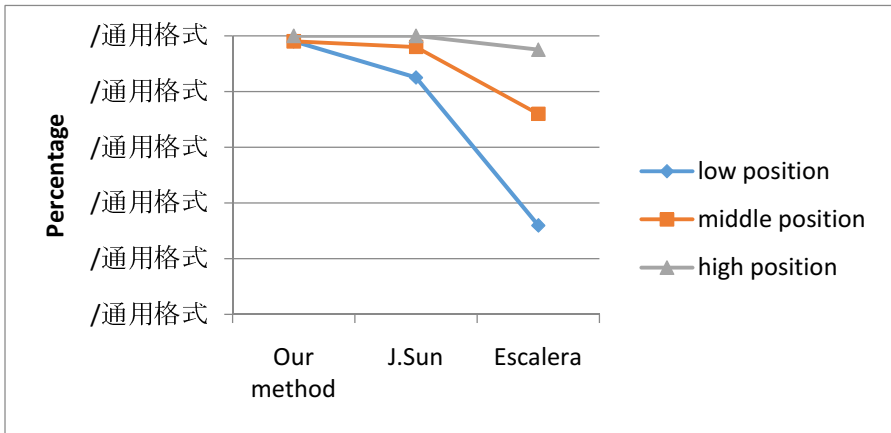


Fig. 15. Chess-board recognition success rates

4.2 Other Factors

We conduct this experiment on 200 images taken by us with different illumination, image sizes and camera poses and another 200 images taken by volunteers with little knowledge about our algorithm except the suggestion of capturing the object and the board at the same time. As shown in **Fig. 16**, we successfully complete calibration of

almost all cameras of the first set, while we fail more over the second set due to some ill-captured images. We think this experiment is very necessary since our algorithm is to serve the users who might care about the performance only. It also helps us to find out the limitations and problems, or at least to make the rules of capturing clearer to users.

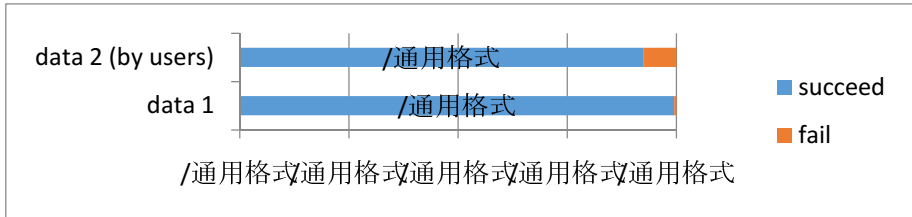


Fig. 16. Experiments on user-captured data

We show some results of chess-board vertices detection and the 3D models correspondingly reconstructed. With the help of our method, we can reconstruct quite accurate and vivid models. (We choose 4 images from each scenario)

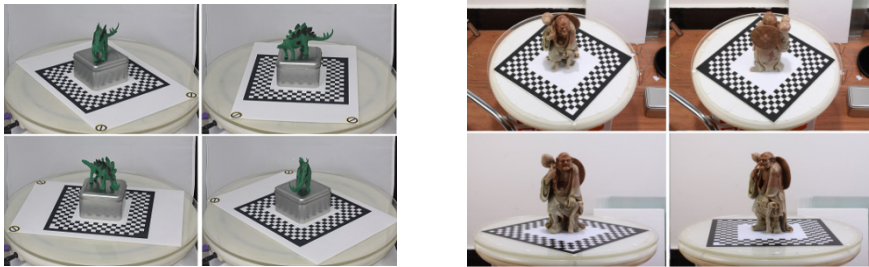


Fig. 17. Original images



Fig. 18. Vertices detection results



Fig. 19. Reconstructed models

5 Discussion and Conclusion

In this paper we justified several properties of the chess-board vertices and designed a multi-scale detector with which we develop an accurate and robust chess-board corner detection algorithm that, according to experiments, performs well even in ill-illuminated and ill-focused conditions. The accuracy and robustness of our detector allows it to be employed in more applications related to chess-board detection. It can also help to refine the vertices detected by other methods.

Acknowledgements. This research was supported by Grant No 2010CB328002 from The National Basic Research Program of China(973 Program), Grant No. 61232014,61121002,61173080 from National Natural Science Foundation of China. Also was supported by Grant No. 2013BAK03B07 from The National Key Technology Research and Development Program of China.

References

1. Moravec, H.: Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover. Tech. Report CMU-RI-TR-3 (1980)
2. Harris, C., Stephens, M.: A Combined Corner and Edge Detector. In: Proceedings of 4th Alvey Vision Conference, pp. 141–151 (1988)
3. Mokhtarian, F., Suomela, R.: Robust Image Corner Detection Through Curvature Scale Space. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(12), 1376–1381 (1998)
4. Canny, J.F.: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 8, 679–698 (1986)
5. Rosten, E., Drummond, T.: Fusing Points and Lines for High Performance Tracking. *IEEE International Conference on Computer Vision* 2, 1508–1515 (2005)
6. Escalera, A., Armingol, J.M.: Automatic chessboard detection for intrinsic and extrinsic camera parameter calibration. *Sensors* 10(3), 2027–2044 (2010)
7. Yu, C., Peng, Q.: Robust recognition of checkerboard pattern for camera calibration. *Optical Engineering* 45(9), 093201 (2006)

8. Sun, W., Yang, X., Xiao, S., Hu, W.: Robust checkerboard recognition for efficient non-planar geometry registration in projector camera systems. In: Proceedings of the 5th ACM/IEEE International Workshop on Projector Camera Systems. PROCAMS 2008, pp. 1–7. ACM, New York (2008)
9. Cai, S., Zang, Z.: A Deformed Chessboard Pattern for Automatic Camera Calibration International Conference on Advanced ICT (2013)
10. Bennett, S., Lasenby, J.: ChESS – Quick and Robust Detection of Chess-board Features. CoRR (2013)
11. Vezhnevets, V.: OpenCV Calibration Object Detection, <http://graphics.cs.msu.ru/en/research/calibration/opencv.html>
12. Bouguet, J., Strobl, Y., Sepp, K., Paredes, W., Arbter, C.: Camera Calibration Toolbox for Matlab (2008), <http://www.vision.caltech.edu/bouguetj>
13. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: SURF: Speeded Up Robust Features. Computer Vision and Image Understanding (CVIU) 110(3), 346–359 (2008)
14. Duda, R.O., Hart, P.E.: Use of the Hough Transformation to Detect Lines and Curves in Pictures. Artificial Intelligence Center (SRI International) (1971)
15. Chen, Y., Sun, J., Wang, G.: Minimizing geometric distance by iterative linear optimization. In: ICPR, pp. 1–4 (2010)
16. Sun, J.: Design and Optimization of a Calibration Point Extraction Algorithm in the Context of 3D Model Reconstruction (2012)

A Mixed-Method Approach for Rapid Modeling of the Virtual Scene Building

Pu Ren¹, Wenjian Wang^{1,*}, Mingquan Zhou² and Chongbin Xu²

¹ School of Computer and Information Technology, Shanxi University, Taiyuan, China
rp_apei@163.com, wjwang@sxu.edu.cn

² School of Information Science and Technology, Beijing Normal University, Beijing, China
mqzhou@bnu.edu.cn, sear2005@163.com

Abstract. Virtual scene building is regarded as a significant part in cultural heritage digital presentation processes. Geometric modeling method, which has been widely used in virtual scene reconstructing, requires lots of manual operations and a long modeling period. This article proposed an approach combining three modeling methods to reconstruct 3D models. Various methods should be adopted regarding to different complexity levels of objects to make the design process more efficient and realistic. In this article, the proposed strategy will be illustrated by presenting a virtual scene of Qiao's Grand Courtyard, one of the most famous ancient residential building groups in Shanxi province. As a result, a simple, efficient, cost-effective and high standard modeling approach has been obtained for the historic heritage digital presentation.

Keywords: Virtual Scene Building, 3D reconstruction, Virtual roaming.

1 Introduction

Possessing a long civilized history, China contributes a precious cultural wealth for the world. Some of these cultural heritages are stored in museums, while others are handed down from generation to generation and rooted in the spirit and blood of Chinese people. With the development of human civilization and technology, digital technology provides a new way in cultural heritage preservation, and the issue of digital preservation of cultural heritage has become an important topic in cross-disciplinary fields, in which digitization of cultural heritage plays a significant role [1].

Virtual scene building, as an application of VR, is an important step in digital display design of cultural heritage [2, 3]. On one hand, the realistic sensation and real-time natural interaction depend on the improvement of the performance of graphic workstation and implications such as digital helmets, data gloves etc. On the other hand, it is more important to focus on the continuous improvement of construction and display techniques in virtual scene and natural interaction. Realistic rendering for each model's structure, function and detail is the essential requirement of an interactive virtual display project.








2 Scene Building in Virtual Interactive Presentation

Nowadays virtual scene modeling in interactive presentation design is mostly carried out by software like 3DS MAX, AUTO CAD, MAYA, etc. The advantage by using this method is that geometric shapes can be controlled by the designer to produce high-quality models with low complexity. However, all the processes totally rely on manual operations, which requires a long modeling period and present a poor sense of reality. Image-based 3D reconstruction [4] can make up these weaknesses well by acquiring three-dimensional structure from the real world directly so as to obtain high geometric accuracy, photo-realism of the results and the modeling of the complete details in a short time with a more portable and flexible approach.

According to various complexities, immovable cultural heritage and artifacts involved in this paper can be divided into three categories:

- Architectural objects with simple shape and flat surfaces. Such as houses, grounds, stone steps, horizontal inscribed boards. In this category, modeling can be achieved by simple geometric transformation while the shape always presents repeatability in some certain directions. Especially in Chinese traditional architecture, compound structures in one building group present to be largely identical with minor differences. Houses in one group are all with similar symmetry and façade. As a result, models in one virtual scene can be adopted with little changes in their texture maps.
- Sculpture objects with a complex shape and concave-convex surfaces. Such as the stone lions, carved drum-like stones and other stone carvings with vivid, flexible, varying shapes. Different from modern factories' mass production, ancient stone sculptures always manufactured by craftsman only one at a time containing their inspiration and cultural emotion. Too much time and energy will be cost in reconstructing this kind of artifacts by traditional geometric modeling and the final effect is often limited by modeler's capacity which can easily reduce the realistic feeling of the whole virtual scene. Therefore, approach of image-based feature points matching [5] should be adopted.
- Artifacts carved in relief with a simple integral shape and complex detail expression. For example, brick and wood carvings are both main forms of traditional relief sculpture technics. Brick carving refers to carving some patterns describing human figures, landscape, flowers and other traditional literary quotations on blue bricks used in building ridges, screen walls, arches over gateway and so forth. These objects often replaced by some ordinary geometric structure with their frontal photograph as mapping texture in traditional geometric modeling, which builds a realistic view for users who are roaming in the scene at a long distance from the model. But when users get very close to these geometric models, flat surfaces without depth information cannot provide stereoscopic impression. So a patch-based multi-view stereo approach will be used in this article to reconstruct a screen wall in Qiao's Grand Courtyard automatically outperforming than traditional method in many ways.

Table 1. Three types of immovable cultural heritage and artifacts

Classification	Examples	Photos	Shortcomings of geometric modeling	Method used in this paper
Architectural objects with a simple shape and flat surfaces	Houses, grounds, stone steps, horizontal inscribed boards	  	—	3D modeling based on modeling software
Sculpture objects with a complex shape and concave-convex surfaces	Stone lions, carved drum-like stones	 	Large project amount; Long modeling period; Poor realistic	3D reconstruction based on image feature points matching
Artifacts carved in relief with a simple integral shape and complex detail expression	Brick carvings, wood carvings	 	Lack of detail information; Poor stereoscopic impression	Patch-based Multi-view Stereopsis Algorithm

3 The Mixed-Method Approach for Rapid Modeling

A mixed-method combining three different approaches is adopted regarding to different complexity levels to make the design process more efficient realistic. The following part will introduce the three modeling approaches by the reconstruction of the Qiao’s Grand Courtyard.

3.1 3D Modeling Based on Modeling Software

The powerful software 3DS MAX is used for reconstruction of majority of artifacts in a virtual scene. The key technology in this process [6] is to optimize the models and scene, including reduction the model and face without bringing down the realistic

feeling. Using texture mapping repeatedly to implement reflection, refraction, concave-convex, transparency and other effects, can paint the model surface more exquisite and realistic(See Fig.1). At the same time minimizing the face number in scene can improve the real-time rendering speed in virtual wandering. Unfortunately when the user reaches a very short distance to the object, maps instead of detail models look too flat than our real world.



Fig. 1. Repeated use of Cutouts to express details

3.2 Reconstruction Based on Image Feature Points Matching

As for the complex-shaped sculptures in this courtyard, the method, which based on matching image feature points, can be adopted to obtain a lifelike effect with a high speed. The reconstruction of space feature points can be done first by computing the position according to its projections on different imaging planes. The object's geometric model under three-dimensional space can be rebuilt based on these 3D points. Take a stone lion as an example, the approach can be decomposed into four steps: firstly, use a hand-held camera to capture the appearance of the sculpture of interest from a number of different overlapping views; secondly, compute camera parameters and 3D coordinates of the matched points by assigning their exact position in different photos; thirdly, rebuild enough 3D feature points of stone lion; and finally link these points in a correct sequence to rebuild a mesh topological structure.

Let C_1 and C_2 respectively denote the local coordinate systems from two camera positions, the corresponding transformation parameters are R_1 , t_1 , R_2 and t_2 . For a point P with the coordinate $X_w(x_w, y_w, z_w)$ in the world coordinate system, its coordinate is $X_{c1}(x_{c1}, y_{c1}, z_{c1})$ in C_1 , and $X_{c2}(x_{c2}, y_{c2}, z_{c2})$ in C_2 , which allows us to write

$$\begin{cases} X_{C1} = R_1 X_w + t_1 \\ X_{C2} = R_2 X_w + t_2 \end{cases} \quad (1)$$

This can be rewrite as the transformation relation between C_1 and C_2

$$X_{C1} = R_1 R_2^{-1} X_{C2} + t_1 - R_1 R_2^{-1} t_2 = R X_{C2} + t \quad (2)$$

Once the positions in two photos of point P are confirmed, we can obtain the relative spatial relation between the two coordinate systems by the equation above.

By repeatedly using this principle we can find out multiple camera spatial position relations between each other.

Put some identification points on the reference substance or the sculpture itself in software called Imagemodeler to complete the calibration of camera [7]. These points are chosen as space restriction points to compute the camera's spatial position. Afterwards, the same point would be located in different views just as Fig.2 shows, based on which every assigned feature point in photos can be located in three-dimensional space.

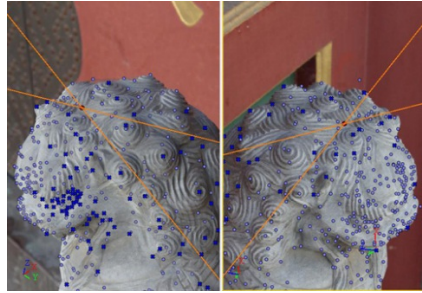


Fig. 2. Determine space position of feature points by two views

Because of the complexity of stone lion's shape, thousands of feature points are needed to describe its curved surfaces. After identifying all the feature points, triangular meshes can be built by linking all these points according to the topological relationship between them. After mapping the texture information from color photos onto 3D mesh surfaces, a model of stone lion is finished as showed in Fig.3.

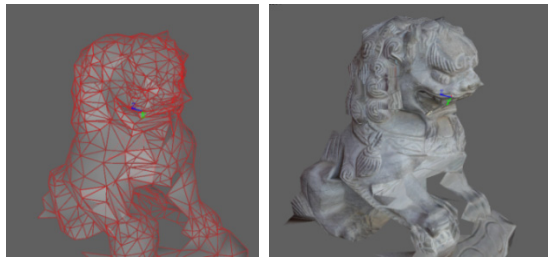


Fig. 3. Left: Stone lion's surface mesh based on feature points Right: Surface model with texture

3.3 Multi-view Stereopsis Algorithm Based on Photos

The 3D reconstruction based on feature points above can obtain complex-shaped models, while in the process of feature points selecting a large amount of human interaction is needed rather than totally automation. As for those artifacts carved in relief with a simple integral shape and complex detail expression, the two approaches talked above cannot generate realistic and accurate models. Furukawa et al. [8] have proposed an algorithm for multi-view stereopsis that outputs a dense set of small rectangular patches covering the surfaces visible in the images (PMVS). This algorithm has

been estimated as one of the state-of-the-art MVS algorithm [9] as automatically generating an accurate, dense and robust result. Poisson Surface Reconstruction proposed in [10] will be used as a post-processing step to convert the set of oriented points produced by PMVS into a triangulated mesh model. Fig.4 shows an effect picture of a screen wall in the Qiao's Grand Courtyard reconstructed by this method.



Fig. 4. Effect of reconstructed models in the system

Multi-view stereo (MVS) matching and reconstruction is a key ingredient in the automated acquisition of geometric object and scene models from multiple photographs, a process known as image-based modeling or 3D photography. Here we utilize a classic algorithm for multi-view stereopsis that outputs a dense set of small rectangular patches covering the surfaces visible in the images.

Stereopsis is implemented as a match, expand, and filter procedure, starting from a sparse set of matched key points, and repeatedly expanding these before using visibility constraints to filter away false matches. The keys to the performance of the proposed algorithm are effective techniques for enforcing local photometric consistency and global visibility constraints.

A patch p is essentially a local tangent plane approximation of a surface. Its center and normal respectively denoted as $c(p)$ and $n(p)$. A reference image $R(p)$ is the pictures used in the matching step. A patch is a 3D rectangle, which is oriented so that one of its edges is parallel to the x-axis of the reference camera (the camera associated with $R(p)$). The extent of the rectangle is chosen so that the smallest axis-aligned square in $R(p)$ containing its image projection is of size $\mu \times \mu$ pixels in size. Let $V(p)$ denote a set of images in which p is visible, I can refer to any images, $V^*(p)$ is the image set ignoring images with bad photometric discrepancy scores by simply adding a threshold.

$$V(p) = \{I \mid n(p) \cdot \frac{\overline{c(p)O(I)}}{|c(p)O(I)|} > \cos(\tau)\} \quad (3)$$

$$V^*(p) = \{I \mid I \in V(p), h(p, I, R(p)) \leq \alpha\} \quad (4)$$

The photometric discrepancy function $g(p)$ for p on $V(p)$ and $V^*(p)$ is defined as

$$g(p) = \frac{1}{|V(p) \setminus R(p)|} \sum_{I \in V(p) \setminus R(p)} h(p, I, R(p)) \tag{5}$$

$$g^*(p) = \frac{1}{|V^*(p) \setminus R(p)|} \sum_{I \in V^*(p) \setminus R(p)} h(p, I, R(p)) \tag{6}$$

where $h(p, \mathbf{I}, \mathbf{R}(p))$ is defined to be a pairwise photometric discrepancy function between \mathbf{I} and $\mathbf{R}(p)$. Given the image pair (I_1, I_2) , $h(p, I_1, I_2)$ equals to 1 minus its NCC (Normalized cross-correlation) value. Patch p will be rebuilt by two steps: (1) Initializes the corresponding parameters: center $c(p)$; unit normal vector $\mathbf{n}(p)$; visible image set $V^*(p)$; the reference image $\mathbf{R}(p)$; (2) Optimizing $c(p)$ and $\mathbf{n}(p)$.

Simple but effective methods are also proposed to turn the resulting patch model into a mesh which can be further refined by an algorithm that enforces both photometric consistency and regularization constraints. The proposed approach automatically detects and discards outliers and obstacles, and does not require any initialization. Fig.5 shows that the whole process is started from a set of disorderly photos and produced a dense points cloud.

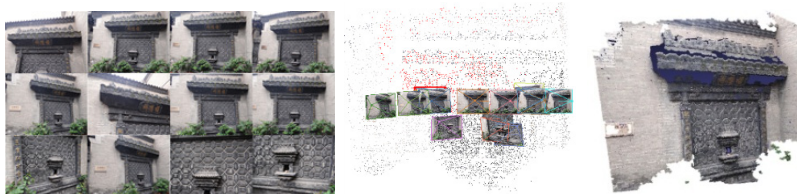


Fig. 5. Reconstruction steps of PMVS

The reconstructed patches form an oriented point model. Despite the growing popularity of this type of models in the computer graphics community, it remains desirable to turn the collection of patches into surface meshes for image based modeling applications in this paper. This process can be cast as a spatial Poisson problem [10], which considers all the points at once, without resorting to heuristic spatial partitioning or blending, and is therefore highly resilient to data noise. Thus, as shown in Fig.6, Poisson reconstruction creates very smooth surfaces that robustly approximate noisy data and fill up the loopholes in point model.

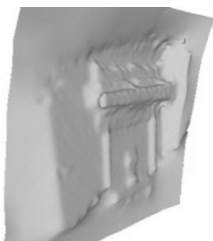


Fig. 6. Result of Poisson Surface Reconstruction

To better visualize the colored model we need a parameterization of the mesh and create the texture from color information of original point cloud. Finally a model of surfaces made entirely of right triangles is generated just as Fig.7 shows.



Fig. 7. PMVS-based Mesh model and texture

4 Results and Discussion

Models created by the first method are most simplified which guarantees a high running speed. But it requires a long modeling period and produces a relatively poor sense of reality. The whole process totally relies on the modeler's manual operation. The second approach based on matching image feature points can reconstruct a mesh model of artifact with complicated shape, while the process of selecting feature points requires some degree of interaction. The third method based on PMVS algorithm can obtain a complete model containing detail information with an entirely automatic procedure.

According to the experiment results, PMVS algorithm based on photos outperforms the geometric modeling method in presenting details. The right-side screen wall in Fig.8 was produced by PMVS algorithm with an obvious detail model in the middle part, while there is just a flat surface in the left-side one.



Fig. 8. Experiment results comparison

(Left: A flat surface based on modeling software; Right: Detail surfaces based on PMVS)

Conclusion can be made by the comparison of the last two methods based on same photo series of stone lion: they both obtain photo-realism of the results with short modeling periods. The difference is the mesh number. The stone lion model produced by the second method has 1,900 meshes, while the model produced by the third method has 64,785 meshes (See Fig.9). As each frame was real-time computed by the video card and CPU in a virtual interactive presentation program, excessive number

of surface in a scene may result in a sharp speed drop of system. Therefore the PMVS algorithm is not suitable to reconstruct all models in the scene, in spite of its accuracy, photo-realism and details presentation.

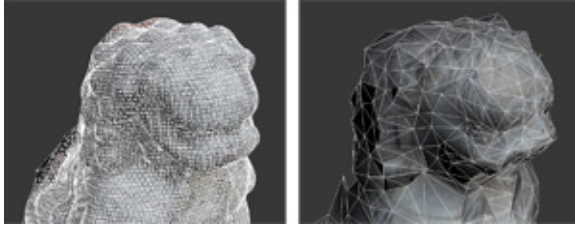


Fig. 9. Experiment results comparison of a stone lion based on photos
(Left: The result of PMVS with 64,785 meshes;
Right: The result of feature points matching with 1,900 meshes)

Adopting the three methods above for different kinds of objects can effectively save the modeling time, guarantee the system to run smoothly and obtain a more realistic effect. Fig.10 shows a panorama of the Qiao's Grand Courtyard which is reconstructed in our experiment including 6 compounds with more than 500 models and 89,200 triangular meshes. Interactive design is implemented in VRPlatform and a virtual roaming system we realized can run smoothly and vividly.



Fig. 10. Panorama of virtual roaming system of Qiao's Grand Courtyard with 89,200 meshes

5 Conclusion

This paper proposed a photo-based strategy which combines three methods to reconstruct different types of immovable cultural heritage. Depending on the different characteristics and complexities of the surface, immovable cultural heritage and artifacts involved in this paper had been innovatively distinguished into three categories. Compared with the traditional modeling methods, this mixed strategy reduces the manual operations, shortens the period of modeling and obtains a more realistic effect. Comparison experiments on the same objects illustrate that each method has its own relative merits and scope of application. Implementation of the virtual roaming system of the Qiao's Grand Courtyard demonstrated that this strategy can be applied

in the use of cultural heritage digital interactive presentation. This article provides an economical, efficient and superior approach for cultural heritage interactive presentation designing, which has a promising prospect in engineering application.

Acknowledgements. The authors would like to thank the anonymous reviewers. This work is supported by the National Key Technology Research and Development Program of China (No. 2012BAH33F04).

References

1. Lynch, C.: Digital collections, digital libraries & the digitization of cultural heritage information. *Microform & Imaging Review* 31(4), 131–145 (2002)
2. Chen, Z.-G., Geng, T., Zhang, Y.: Introduction on product interactive display design. *Art and Literature for the Masses* 18, 074 (2011)
3. Huang, Y.-X., Tan, G.-X.: The research on digital protection and development of Chinese intangible cultural heritage. *Journal of Huazhong Normal University (Humanities and Social Sciences)* 51(2), 49–55 (2012)
4. Remondino, F., El-Hakim, S.: Image-based 3D Modelling: A Review. *The Photogrammetric Record* 21(115), 269–291 (2006)
5. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision* 47(1-3), 7–42 (2002)
6. Ren, P., Wang, W.-J., Bai, X.-F.: Reconstruction of Imperial University of Shanxi Based on Virtual Reality Technology. *Computer Simulation* 29(11), 20–23 (2012)
7. Qiao, J., Guo, J.-H., Lan, T.-L.: Three-dimensional Image-based Modeling for Reconstruction of Cultural Relics. *Sciences of Conservation and Archaeology* 23(1), 68–71 (2011)
8. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(8), 1362–1376 (2010)
9. Shi, L.-M., Guo, F.-S., Hu, Z.-Y.: An Improved PMVS through Scene Geometric Information. *Acta Automatica Sinica* 37(5), 560–568 (2011)
10. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: *Proceedings of the Fourth Eurographics Symposium on Geometry Processing* (2006)

An Improved Method of Building Rapid 3D Modeling Based on Digital Photogrammetric Technique

Zimin Zhang, Ying Zhou, Jian Cui, and Hao Liu

Department of Civil Engineering, Shandong Jianzhu University,
Shandong, 250101, China

Abstract. Building 3D modeling is a fundamental but expensive works to digital city engineer. For cutting off the workload of that, an improved rapid modeling method based on digital photogrammetric technique is proposed. Two main issues are resolved in the method. The first is how to trace and recognize the roof surfaces of a building according to its profile lines, and the second is how to create right solid models for complicated buildings usually with multiple parts. The paper gives a detailed description about the solutions and involved algorithms. Finally, multiple buildings with different roof styles are selected to test the improved method. Results show that it can get right 3D models for common buildings, and reduce the workload of delineating roof lines and possible model mistakes.

Keywords: Rapid modeling, Building, 3D modeling, Digital Photogrammetry, Digital city.

1 Introduction

The establishment of city's 3D Model is the foundation of digital city that is widely carried out in China at present, and buildings constitute the main body of the model. Therefore, building 3D modeling has become an important task for urban surveying and mapping bureau of China.

Building 3D modeling consists of two main steps: making the structure models of buildings and pasting the texture pictures for building surfaces. The structure model is usually name "white model" for its surfaces assigned with a uniform color. After the pasting of true texture pictures on the surfaces by image cutting, correcting and mapping, a realistic building 3D model is finished. Presently, the process is primarily done by manual works, and consequently demand high costs. A introduce of rapid modeling methods and computer systems to improve the efficiency of that is significant.

At present, rapid modeling systems are immature. Some methods proposed by researchers include: 1) using Light Detection and Ranging (LIDAR) point cloud data to extract the outlines of buildings and make a wireframe model[1]. 2) Constructing building model with the parametric information extracted from images by morphological scale space image processing [2]. 3) Using tilt image to extract building's texture in four directions [3]. This paper presents a method utilizing structural vector line extracted by digital photogrammetric technique to generate building models automatically.

2 Building Rapid Modeling Method

2.1 Process Design

With the aid of the digital photogrammetric system, the profile lines of the top of buildings and their ground Digital Elevation Model (DEM) can be extracted from stereo images. Then the structure models are created automatically according to the two datasets. Two issues need to solve in the process, which are how to recognize the top surfaces from the out-of-order lines, and generate the right structure solid models to the complicated buildings usually comprised of multiple related parts or containing one or more holes. Although some digital photogrammetric workstations also have the capability of automatically generating the structure models, additional restrictions are usually added to the process, such as profile lines for a surface must be collected separately and recorded explicitly, which will lead to duplicated delineation of the outlines shared by multiple surfaces, and therefore possible disagreement of the collected lines.

To solve the problem, an improved process is designed and implemented, which contains some algorithms for tracing and recognizing the structure surfaces and constructing the solid models for complicated buildings. As shown in Figure 1, the new process includes three major steps, the recognition of structure surfaces, the creation of solid models, and the reconstruction of models.

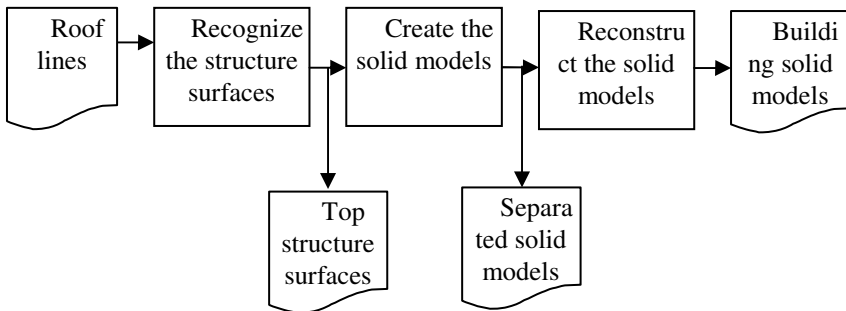


Fig. 1. The improved building 3D modeling process

2.2 Recognition of Building Structure Surfaces

The process traces the building roof lines extracted by the digital photogrammetric technique and recognizes the roof surfaces. The roof surface is a closed and coplanar three-dimensional polygon. The roof lines can be three-dimensional lines, polylines or polygons. This process can be divided into the following three steps.

(1) Extract the relevant roof lines

Relevant roof lines refer to the lines contacting or intersecting with each other in the 3D space. The extraction step starts with a random seed line specified, and

a recursive depth priority search algorithm is developed to achieve all related lines. Figure 2 shows the process to extract all the relevant lines related to a seed line marked as red color.

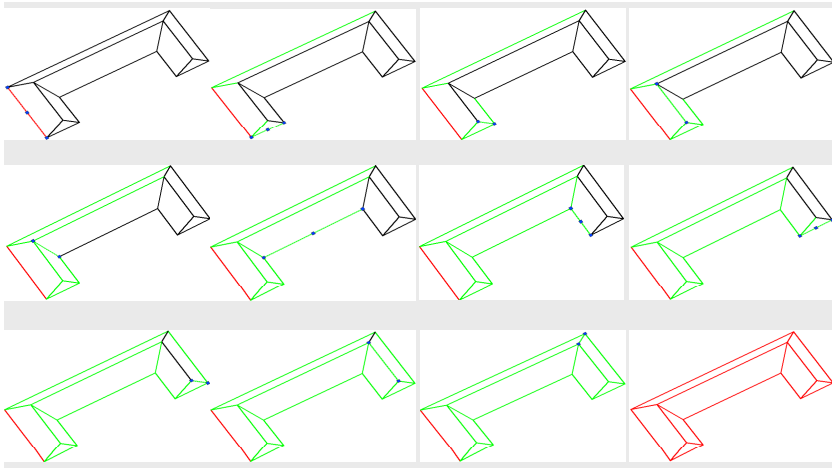
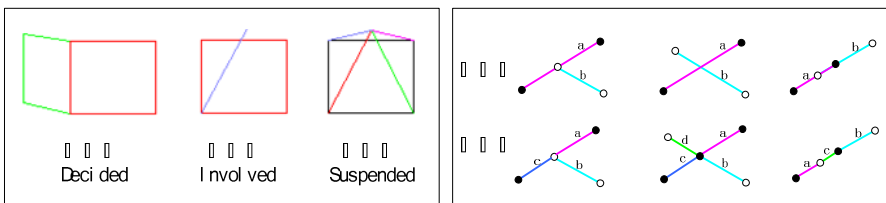


Fig. 2. The extraction progress of relevant lines

(2) Reconstruct the relevant lines

For the recognition of roof surfaces, the abstracted relevant lines need to be reconstructed. First of all, the relevant lines are divided into three categories according to whether they can form roof surfaces directly: decided, involved and suspended lines. The decided lines can constitute roof surfaces independently, and also are not utilized by other surfaces. The involved lines also can shape surfaces, but maybe utilized by other surfaces, i.e. interacting with other relevant lines. The suspended lines cannot form surfaces itself while may be possible with other lines together. The red lines in Figure 3a indicate the three line types.

Secondly, the decided lines are extracted directly to build surfaces, while the others are broken at intersections to form simple line segments. Figure 3b shows the results of breaking the lines.



a) Three line types

b) The results of breaking the lines

Fig. 3. Line types and its reconstruction

(3) Recognize the top surfaces

Firstly, a directed graph expressing the connecting topology of the reconstructed lines is created. As demonstrated by Figure 4, the lines are delegated by the edges of the graph, and the nodes are represented as the vertices of the graph, which also stores a chain set to record all the edges connected to it.

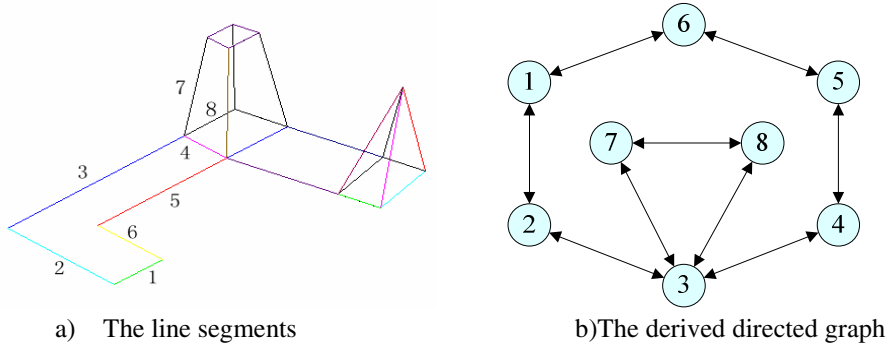


Fig. 4. The creation of the directed graph

Secondly, a recursive depth priority search algorithm is developed to extract the roof surfaces according to the directed graph. Given a random edge, the algorithm searches other edges along one direction and records coplanar edges continuously until a closed surface is formed or no surface is achieved. After all the edges are processed, all the roof surfaces are found.

At the end, the roof surfaces are sorted by counter clockwise and the vertical ones are eliminated for satisfying the requirements of creating solid models.

2.3 Create Building Solid Models

The building solid models can be created utilizing the identified roof surfaces and DEM generated by the digital photogrammetric technique. The process consists of two main steps as following.

(1) Query the ground elevation

Each roof surface is projected onto the ground vertically and the DEM is queried inside the projected region by a specified resolution. Consequently, a series of ground elevation values are obtained, and there the minimum value is considered as the building ground elevation.

(2) Creating the solid models

The solid models are created by stretching the roof surfaces down to the ground elevation. While at this moment the solid models constituting a building are separated, some further processing is still needed to get the right model. Figure 5 is an example to create solid models by stretching.

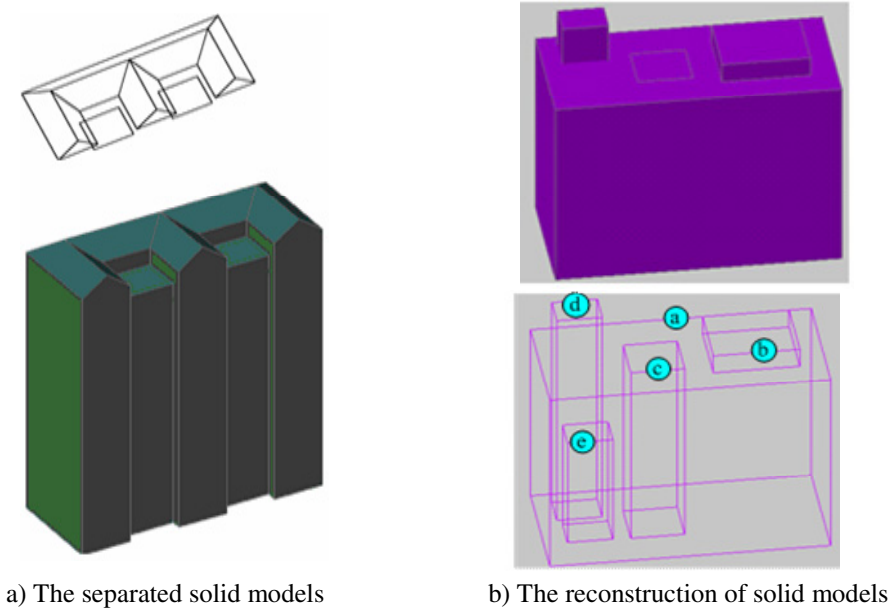


Fig. 5. The creation and reconstruction of building solid models

2.4 Reconstruct Building Solid Models

The reconstructing process combines the separated solid models into a complete one. To the complicated buildings, solid union operation usually can't receive right results, and some more combining rules need to be designed. The following is the rules gotten by analyzing the different building structures.

Firstly, all solid models of a building are sorted by the top elevation in descending order. The later combining operations will be fulfilled according to that order.

Secondly, the following rules are taken when combining two solid models, which are formulated based on the interference (intersection) relationship of them:

- (1) Union the two models when they interfere but don't contain.
- (2) The model with higher top elevation subtracts the other one when they interfere and contain.

An example is shown in the figure 5b. In the five solid models, three pairs interfere but don't contain, i.e. (a, c), (a, d) and (a, e), and two pairs interfere and contain, i.e. (a, c) and (a, e).

By the above rules, the building solid models are created completely. After that, the bottom surfaces need to be deleted for the building models are usually placed onto a landscape layer composed of DEM and DOM.

3 Experiments and Verification

According to the designed process, a rapid building modeling system is developed. Then various types of buildings are selected to validate the proposed method, including ordinary flat roof, slope roof, spire, island-style, ordinary connected building and complicated one with corridors, and arch building. Figure 6 shows the structure lines collected and the 3D models created by the system.

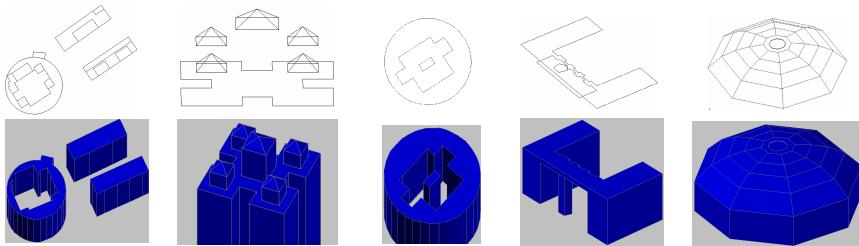


Fig. 6. The delineated structure lines and created solid models

By these testing cases, it is confirmed that the proposed method can create right solid models for varieties of common buildings. Shared lines are needed to delineate only once, and hence possible disagreements are avoided. As a result, the efficiency of building 3D modelling is improved.

4 Conclusions

Building 3D modelling has become an important fundamental task of digital city engineer. This research presents an improved building rapid modelling method based on the structure lines extracted by digital photogrammetric technique, which can avoid duplicated delineation of shared structure lines. By the testing of various types of buildings, the proposed method can successfully complete the creation of solid models for most common buildings, and give great benefit to promote the efficiency of building 3D modelling.

Acknowledgements. The research work was supported by Shandong Province Natural Science Foundation under Grant No. ZR2011EEQ006.

References

1. Li, Y., Feng, Z., Wang, H.: Three-dimensional modeling of buildings based on LIDER point cloud. *Forest Inventory and Planning* 36(6), 29–31 (2011)
2. Wei, Z., Sun, Y., Ji, X., Yang, M.: Rectangular building auto extraction of digital city. *Computer Science* 36(1), 211–215 (2009)
3. Gui, D., Lin, Z., Zhang, C.: Research on construction of 3D building based on oblique images from UAV. *Science of Surveying and Mapping* 37(4), 1–8 (2012)

3-D Reconstruction of Three Views Based on Manifold Study

Li Cong¹, Zhao Hongrui¹, Fu Gang¹, and Peng Xingang²

¹ Institute of Geomatics, Department of Civil engineering, Tsinghua University, Beijing, China

² College of Geoscience and Surveying Engineering, China University of Mining & Technology, Beijing, China

Abstract. Obtaining 3-D reconstruction directly and expediently for the real world has become a hot topic in many fields. A 3-D reconstruction method of three views based on manifold study is proposed. Firstly, the fundamental matrix is estimated by adjacent view and optimized under three views constraint. Then 3-D point cloud is reconstructed after getting the projection matrixes of views. Further more, benefitting from minimum spanning tree, outliers are almost excluded. To increase point cloud's density, the optimized 3-D point cloud is interpolated based on Radial Basis Function. Afterwards, the dense point cloud is mapped to two dimensional plane using manifold study algorithm, and then divided into plane Delaunay triangle nets. Completing that, the topological relations of points are mapped back to 3-D space and 3-D reconstruction is realized. Many experiments show the method proposed in paper can achieve 3-D reconstruction for three views with quite good results.

Keywords: Manifold study, three views, 3-D reconstruction, fundamental matrix, minimum spanning tree.

1 Introduction

3-D reconstruction is a fundamental issue in the fields of computer vision and photogrammetry. Approaches based on image sequences can directly and quickly reconstruct for the real world just rely on epipolar geometry of adjacent view. Due to its low cost and immediately color acquisition power, image-based 3-D reconstruction has broad application prospects and becomes a hot topic in many fields.

The fundamental matrix is the algebraic representation of epipolar geometry, and it's the only geometry information obtaining from un-calibrated image sequences. So accurate computation of fundamental matrix is an necessary and important step to realize 3-D reconstruction. Fundamental matrix is commonly calculated by matching points between views, and its methods are mainly divided into 3 categories: linear algorithms, non-linear algorithms and robust algorithms[1]. Because of the ability of distinguishing mismatches, robust algorithms could calculate correct fundamental matrix by selecting inliers set. Consequently, a vast amount of research effort has been dedicated to robust methods. However, not only the traditional robust methods[2] but also the advanced methods[3,4,5,6,7,]based on them just are adopted with the constrain

of epipolar geometry, even the novel methods[8,9,10]relied on optimized model of matching points' residuals. When there exist mismatches, those methods could not remove them absolutely, getting a better inliers set in the case of possible greater probability. The trifocal tensor is a multiple view object in three views, which could be described by the geometry of a point-point-point .So it can behave precise constraint between matching points. With the constraint of three views, paper[11]excludes more mismatches between adjacent view during the process of 3-D reconstruction and achieve better results. In the same way, this article gets more accurate matching points to build inliers set without solution of trifocal tensor. Then fundamental matrix is calculated precisely with the optimal inliers set.

The projection matrix could be obtained through the decomposition of fundamental matrix. Building and solving equations between the matching points and its corresponding projection matrix, the reconstruction of 3-D point cloud is completed. To acquire matching points, SIFT algorithm is adopted. As a result, point cloud mainly represent the steady feature in the view. But, its are sparse and uneven distributed. Considering the characteristics of the reconstructed 3-D point cloud, a new method to reestablish the geometric surface model is also proposed in this article.

Nowadays, how to quickly and accurately reconstruct objects' geometric surface is also a heated question by discrete 3-D points in many fields. Because of the excellent performance in describing object model, TIN(Triangulated Irregular Network) has attracted lots of researches and wide application. The main methods of it could be classified into Delaunay triangulation method and implicit surface fitting method.

On account of the obtaining point cloud is steady, sparse and uneven distributed, methods have its own advantages and disadvantages in 3-D TIN reconstruction. Sculpture algorithm[12] and incremental surface growing algorithm based on Delaunay triangulation could gain high-quality triangular mesh surface. But its complexity of time and space relies on points number. When the 3-D points set is large, its efficiency becomes low. Its also couldn't deal with noisy and uneven distributed data. What's more, it's sensitive to sampling density. Local triangulation method[13]has higher efficiency in recovering TIN, but the topological relation between points is easily to be changed. Making smaller differentiation between reconstructed TIN and object surface, manifold study algorithm[14]guarantees the essential relationship between points. However, it also requires a certain degree of sampling density. Implicit surface fitting method[15,16]could describe shape of complicated object, and it rebuilds watertight surface, being robust to tiny noise. Limits of it is that the reconstructed TIN doesn't through the original 3-D points. So implicit surface fitting method with manifold study method are combined in this article. Firstly, uniform 3-D point cloud is obtained by RBF interpolation and resampling. Then the reconstruction of TIN is achieved by manifold study, which guarantees the correctness of points topological relationship and TIN via original points at the same time. Finally, 3-D reconstruction is completed after texture mapping.

The main contributions of the method proposed in this article are listed as follow:

(1) Considering the weakness constraint of epipolar geometry, this article gets more accurate matching points inliers set instead of using three views constraint, with no

computation of trifocal tensor. Aim to get precise result of fundamental matrix is came true.

- (2) Remove outliers in 3-D point cloud based on the theory of minimum spanning tree.
- (3) Combined implicit surface fitting method with manifold study method to complete reestablish of 3-D TIN, taking the characteristics of point cloud into account, witch is steady, sparse and uneven distributed.

2 Calculation of Fundamental Matrix Based on Three Views

2.1 Three Views Constraint

Three views constraint-the trifocal tensor, has analogous properties to the fundamental matrix of two view geometry: it is independent of scene structure depending only on the relations between the views. The view matrices may be retrieved from the trifocal tensor up to a common projective transformation of 3-space, and the fundamental matrices for view-pairs may be retrieved uniquely.

The essence of the epipolar constraint over two views is that rays back-projected form corresponding matching points are coplanar, a weak geometric constraint, as in figure 1

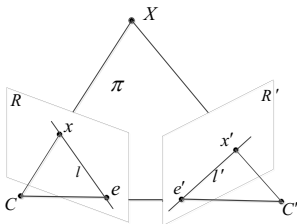


Fig. 1. Epipolar constraint

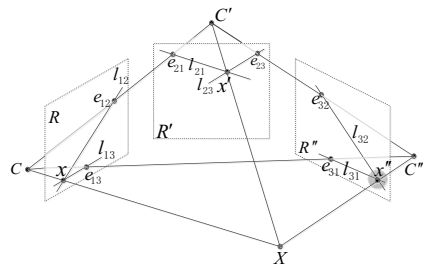


Fig. 2. Geometric constraints of Three-View

C and C' are the centers of the two cameras when they are getting the views R and R' . The camera baseline $C C'$ intersects each image plane at the epipoles e and e' , x and x' are the points which imaged by a 3-space point X in two views R and R' . Any plane π containing the baseline $C C'$ and point X is an epipolar plane, and intersects the image planes in corresponding epipolar lines l and l' . l' is the epipolar line of point x , and it passes through epipoles e' and point x' , the corresponding point of x in another image. So, given a pair of image, it was seen in figure 1, the epipolar geometry constraint can only guarantee that any point x' in the second image matching the point x must lie on the epipolar line l' . But exact position where the point is could not be determined. Compared with epipolar geometry, the three views geometry is the ability to transfer from two views to a third: given a point correspondence over two views the position of the point in the third view is determined as in figure 2.

Symbols in figure 2 are the same with those in figure 1. e_{ij} represent the epipole on the i^{th} image that is imaged from the j^{th} camera center. Similarly, l_{ij} represent the epipolar line on the i^{th} image of the projection point on the j^{th} image. x, x' and x'' are the projection points of 3-space point on the image R, R', R'' . they are correspondence points. So a conclusion can be drawn from three views constraint: any projection point on one image of a 3-space point is the same with intersection of epipolar lines on the very image of its correspondence points.

To distinguish whether a pair of points is matching points or not ,the distance from point to its matching point's epipolar line is always used as the judgment in traditional robust algorithms. However, the exact position of matching points couldn't be determined merely with epipolar geometry constraint. Couldn't it guarantee the matching is right, even though the matching points have tiny residual(distance from point to its matching point's epipolar line). Suppose there is a point X in 3-space, as in figure 3, x is its projection point in image R and x' should be another projection point in image R' . Because of the detection error, x'_a, x'_b and x'_c are the undetermined points of x' . The distance of dotted line in figure 3 represent the magnitude of residual. Comparing the residuals of x'_a and x'_b , it's easily to make sure that x'_a is the right one. Actually, has a smaller residual than x'_a , and it also probability to be regard as the correct matching point . But it is a error obviously shown from the figure 3.

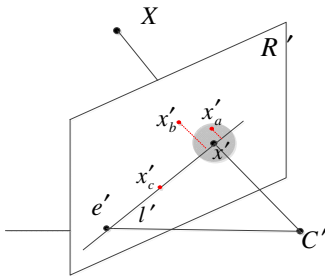


Fig. 3. Correctness judgment of matching points

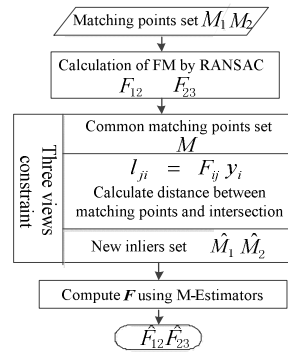


Fig. 4. Flow diagram of fundamental matrix calculation

Three views has stronger constraints between the matching points. For an example as shown in figure 3, x and x' are the matching points, l_{31} and l_{32} are their epipolar lines on image, whose intersection should be the same point with point x'' in theory. Because of the noise produced during the extracting of matching points, the shadow region as shown in figure 3, maybe the area where real point x'' is. But distance between the point of intersection and the real matching point should be limited below a certain threshold. In conclusion: distance between matching point and the intersection of corresponding epipolar lines should be large, when the matching points in three views are not right. Based on that in this article, a more accurate inliers set is obtaining by getting rid of mismatches.

2.2 Calculation of Fundamental Matrix

Benefitting from the feature detecting and matching of SIFT algorithm, matching points set $M_1=\{x_1,x_2\}$ between the first and third view is acquired, as well as set $M_2=\{x_2,x_3\}$ between the second and third view, and set $M=\{x_1,x_2,x_3\}$ among in all of the three views. Then the fundamental matrices F_{12} and F_{23} are estimated by RANSAC algorithm. Calculating epipolar lines of matching points in set M , l_{ij} represents the epipolar line on the i^{th} image of the projection point on the j^{th} image and it is written in Form of homogeneous coordinates, y_i is the homogeneous coordinates of points in the i^{th} view. The next step is to acquire the intersection point of epipolar lines, and the distance between matching point and that intersection also be calculated. Finally, those initial matching points sets M_1 and M_2 are updated by new sets $M_1'=\{x_1,x_2\}$ and $M_2'=\{x_2,x_3\}$.

Using the new matching set, more correct result of fundamental matrix is acquired. To obtain more precise result, M-Estimators algorithm is used after that. The basic theory of M-estimators is to make a guarantee that the probability of being reduced is larger than of producing error for mismatches. By this way, the precision of the computation is improved via reducing the influence of data containing noise. The whole process of calculating fundamental matrix is shown below as figure 4.

3 Optimization and Interpolation of 3-D Point Cloud

3.1 Reconstruction of 3-D Point Cloud

Camera intrinsic parameters could be acquired by camera self-calibration. As intrinsic parameters known, fundamental matrix is upgraded to essential matrix. Taking decomposition of essential matrix, the relative extrinsic parameters between adjacent two views are obtained. What's more, the projection matrices and are recovered as well. Form projection process, equations could be built: $x=PX$ and $x'=P'X$. X are the homogeneous coordinates of a 3-space point. Satisfied with equations above at the same time, simultaneous equations could be draw as equations (1)

$$\begin{pmatrix} P^1 - x_i^1 P^3 \\ P^2 - x_i^2 P^3 \\ P'^1 - x_i'^1 P'^3 \\ P'^2 - x_i'^2 P'^3 \end{pmatrix} X_i = 0 \tag{1}$$

i represents the i^{th} point, x^1 is the first element of point, is the first row of matrix. Projection matrix of the third view also could be calculated by the back calculation of equations (1) for the points ,which have projective points in all three views.

3.2 Optimization of 3-D Point Cloud

There is no ways to avoid existing of noise, even error, during the processes of point detecting and matching. Three point of containing larger noise or error is called outlier in this article. And they have bad influence in TIN reconstruction result. To exclude them, the theory of minimum spanning tree is adopted. Firstly, the minimum spanning tree is built by Prim algorithm according to the distance between the points. So the point cloud could be regard as being connected by many bridges, and the length of bridge is distance between points. As outlier is far away from the target point cloud, larger is the length of the bridge between them. Seen from this view, outlier is like a island floating outside of target point cloud. Consequently, selecting an appropriate length of bridge as a threshold, the point cloud will be divided into many clusters by breaking bridge whose length is bigger than threshold. Maintaining the point cloud cluster owning most points, other clusters is removed as outliers as shown in figure 5

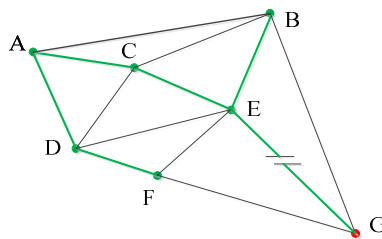


Fig. 5. Sketch of removing outlier

The green bold line is the Prim minimum spanning tree and G represents outlier. For an example, G could be removed when the length of edge EG is selected as a threshold.

3.3 Interpolation of 3-D Point Cloud

When the non-uniform sampling points on manifold are used for dimensionality reduction, its subset mapped in Low dimensional Euclidean space is non-convex, and the mapping result is always not right. Consequently, the topology reconstruction based on manifold study should be taken in the uniform distributed 3-D point cloud. So the interpolation and resampling for original point cloud is necessary to get uniform distribution point cloud. The RBF(Redial Basis Function)based on implicit surface fitting method is used to achieve the interpolation in this article.

RBF is a real-valued function whose value just relies on the distance from the origin, as well as a scalar function which is symmetrical along radial. So the interpolation in 3-D space based on RBF could be described as: a points set $\{X_i, i=1,2,\dots,N\}$ in R^3 is given, the function value for every correspondence point is $\{f_i, i=1,2,\dots,N\}$. A function $F: R^3 \rightarrow R$ is built ,which is satisfied for every sample point

$$F(X_i) = f_i \quad i = 1, 2, \dots, N \tag{2}$$

There are infinite solutions of equation(2). But the ideal result should make the interpolation surface smooth and its energy is as little as possible at the same time. In other word, the real result obtains when the vale of energy function is getting least. So the general form of minimum energy solution can be written as equation (3)

$$F(X) = \sum_{i=1}^N \lambda_i \Phi(|X - X_i|) + P(X) \tag{3}$$

X represents any point in interpolation surface; X_i is a sample point; $|X - X_i|$ is Euclidean norm; Φ is RBF, and λ_i is the weight factor for every correspondence RBF. To guarantee the linearity and continuity for interpolation surface in equation (3), $P(X)$ is defined as follow:

$$P(X) = c_1x + c_2y + c_3z + c_4 \tag{4}$$

c is undetermined coefficient, $\Phi_{ij}=\Phi(|X_i - X_j|)$, Linear system like equation (5) is obtained:

$$\begin{bmatrix} \Phi_{11} & \dots & \Phi_{1N} & x_1 & y_1 & z_1 & 1 & \lambda_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Phi_{N1} & \dots & \Phi_{NN} & x_N & y_N & z_N & 1 & \lambda_N \\ x_1 & \dots & x_N & 0 & 0 & 0 & 0 & c_1 \\ y_1 & \dots & y_N & 0 & 0 & 0 & 0 & c_2 \\ z_1 & \dots & z_N & 0 & 0 & 0 & 0 & c_3 \\ 1 & \dots & 1 & 0 & 0 & 0 & 0 & c_4 \end{bmatrix} = \begin{bmatrix} f_1 \\ \vdots \\ f_N \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{5}$$

Combination coefficient $\{\lambda_1, \dots, \lambda_N\}$ and multinomial coefficient $\{c_1, c_2, c_3, c_4\}$ are calculated by getting resolution of equations (5). Substituting that result into equation, interpolation function $F(X)$ is obtained. Further more, the zero level set of $F(X)$ is the vary interpolation surface of sample points.

4 Topology Reconstruction of TIN Based on Manifold Study

Manifold study is a novel nonlinear dimensionality reduction method. Compared with traditional methods, manifold study will reveal the potential key low dimensions from high dimensionality, and it also has many advantages, such as: adaptive to nonlinear data structure, smaller parameter selection, and readily understood for its construction. Base on manifold study, this article takes a dimensionality reduction on the optimized original 3-D point cloud to generate planar point set. The Delaunay triangulation of the planar point set is a classic problem in computational geometry. For arbitrary planar point set, its Delaunay triangulation existed uniquely. So the error between original point cloud and its low dimensional mapping could be minimized in account of some

constraints. Compared with other dimensionality reduction methods, manifold study approach will preserve the natural connections between data, and make the topological difference between reconstructed triangulation and the real object surface smaller. So we use ISOMAP[17] algorithm one of typical manifold study for the triangulation reconstruction in this paper.

ISOMAP algorithm is proposed based on Multidimensional Scaling Analysis (MDS). It preserves the geometric feature of manifold globally. The idea of ISOMAP is similarity preserving, transforming data from high dimensional space to low dimensional space, preserving the inner geometric relationship (preserving the geodesic distance between two points) between data points. In the MDS, distance matrix is constructed base on Euclidean distances of original data points. What is different in ISOMAP, the distance matrix is constructed base on geodesic distances. Because the geodesic distance is represented by integral form in continuous space, It is impossible be used in discrete computation directly. So the in ISOMAP, shortest path distance is used in ISOMAP for the approximation of the geodesic distance. ISOMAP algorithm is a global method for preserving the geodesic distances for the all data points.

ISOMAP algorithm could be described by three steps as follow:

(1) Construct the neighborhood criterion structure

Firstly, the criterion structure of neighborhood relationship is constructed in this step. Then the adjacency graph is created based on the above mentioned criterion structure. The neighborhood relationship is judged by the Euclidean distances between data points. Suppose the number of data points is N . A $N \times N$ Euclidean distance matrix G_E is constructed: for any two data points x_i and x_j , if the Euclidean distance between them is less than a given threshold ε , or x_j is one k -neighbor of x_i , $G_E(i,j)=Dist_E(x_i, x_j)$ Where $Dist_E(. , .)$ denotes the Euclidean distance between these two points. Otherwise, $G_E(i,j)=\infty$.

(2) Compute the shortest path

This step is used to estimate the geodesic distances between data points. First, a $N \times N$ shortest path matrix G_S is created. The elements in G_S is judged using the corresponding value in G_E : $G_S(i,j)=\infty$, if $G_E(i,j)=\infty$; otherwise $G_S(i,j)=Dist_S(x_i, x_j)$, where $Dist_S(. , .)$ denotes the shortest path distance between these two points. In this step, Dijkstra algorithm can be used for the shortest path calculation.

(3) Compute the low dimensional embedding Y

By using the classical MDS method, the low dimensional embedding Y is calculated based on the shortest path distance matrix (the approximation of geodesic distance matrix). Assume $H=-(I_n-\eta_n\eta_n^T/n)G_S(I_n-\eta_n\eta_n^T/n)/2$, where I_n is a n -order identity matrix, and $\eta_n=[1, \dots, 1]^T \in R^n$. Suppose $\lambda_1, \lambda_2, \dots, \lambda_d$ are the largest d eigenvalues of H , and their corresponding eigenvectors are $\beta_1, \beta_2, \dots, \beta_d$. The final low dimensional embedding $Y=diag(\lambda_1^{1/2}, \lambda_2^{1/2}, \dots, \lambda_d^{1/2})[\beta_1, \beta_2, \dots, \beta_d]^T$.

5 Experiments and Conclusion

The whole process of 3-D reconstruction method proposed in this article is organized as shown in figure 7. And each step of the process is programmed and implemented by ourselves in Matlab and C#. Figures from 8 to 11 are the details and results during the reconstruction process and figure 12 is the final reconstruction result.

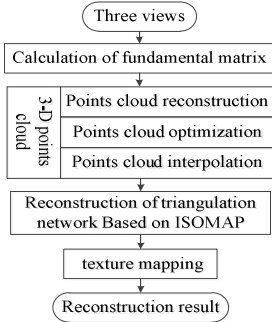


Fig. 6. whole process of 3-D Reconstruction



Fig. 7. Original three views

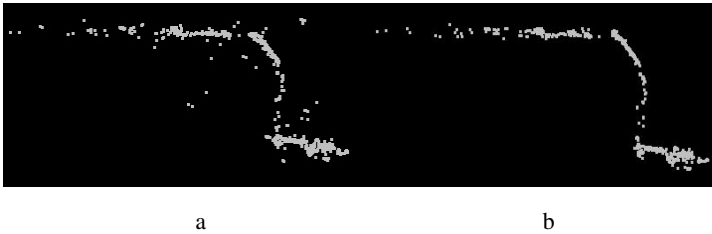


Fig. 8. a. Original 3-D point cloud and b. the optimized point cloud

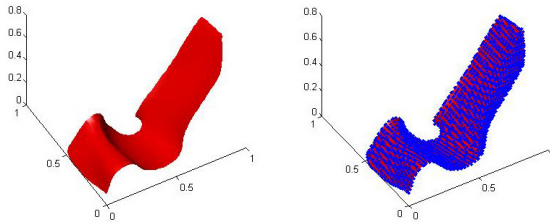


Fig. 9. Interpolation surface of point cloud and its interpolation points

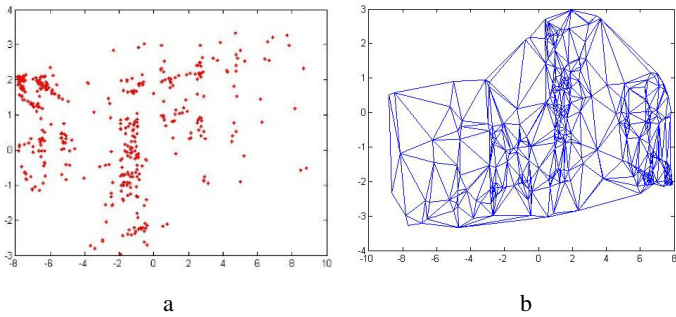


Fig. 10. a. Result of dimensionality reduction using ISOMAP and b. subdivision of plane Delaunay triangle nets

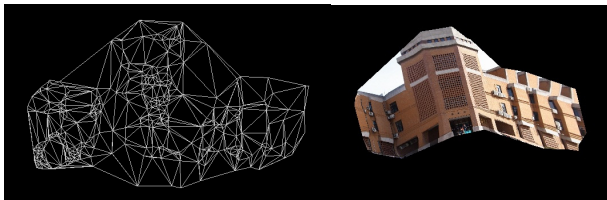


Fig. 11. 3-D TIN and reconstruction result with texture mapping

In conclusion, the 3-D reconstruction method proposed in this article, which based on manifold study by unit of three views, achieves the following optimizations:

- (1) With no computation of trifocal tensor, this methods obtains precise calculation of the fundamental matrix relying on the optimized inliers set, getting rid of mismatches by three views constraint, which could be detected just by epipolar geometry. Meanwhile, more 3-D points are obtaining than two views. It overcomes the insufficient that the number of 3-D points declines rapidly during adding the distance between adjacent views to some extent. Consequently, this method expands the range of 3-D reconstruction applications based on image sequences.
- (2) To exclude the outliers in reconstructed 3-D points, the theory of minimum spanning tree is applied. Then the accurate initial point cloud is obtained, which is the basic of 3-D topological reconstruction.
- (3) The initial 3-D points represent the obvious feature points of structure or texture in actual scene, application of manifold study could make sure that those points must be on the reconstructed scene. what's more, the correctness of topological relations is also be guaranteed.

References

1. Armangué, X., Salvi, J.: Overall view regarding fundamental matrix estimation. *Image and Vision Computing* 21(2), 200–205 (2003)
2. Zhong, X. H.: Research on methods for estimating the fundamental matrix. Jilin University (2005) (基本矩阵计算方法的研究)

3. Torr, P.: Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *International Journal of Computer Vision* 50(1), 35–61 (2002)
4. Carro, A.I., Morros, R.: Promeds: An adaptive robust fundamental matrix estimation approach. In: 3DTV-Conference, The True Vision - Capture, Transmission and Display of 3D Video, pp. 1–4 (2012)
5. Li, Y., Velipasalar, S., Gursoy, M.C.: An improved evolutionary algorithm for fundamental matrix estimation. In: 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 226–231. IEEE Press, Krakow (2013)
6. Shi, X.B., Liu, F., Wang, Y., et al.: A Fundamental Matrix Estimation Algorithm Based on Point Weighting Strategy. In: 2011 International Conference on Virtual Reality and Visualization, Beijing, pp. 24–29 (2011)
7. Calderon, D.B., Maria, T.: An approach for estimating the fundamental matrix. In: 2011 6th Colombian Computing Congress, Manizales, pp. 1–6 (2011)
8. Brandt, S.: Maximum likelihood robust regression with known and unknown residual models. In: Proceedings of the Statistical Methods in Video Processing Workshop, in Conjunction with ECCV, Copenhagen, pp. 97–102 (2002)
9. Brandt, S.: Maximum likelihood robust regression by mixture models. *J. Journal of Mathematical Imaging and Vision*. 25(1), 25–48 (2006)
10. Lu, S., Lei, Y., Kong, W.W., et al.: Fundamental matrix estimation based on probability analysis and sampling consensus. *Control and Decision* 42(2), 425–430 (2012), (基于模糊核聚类的鲁棒性基础矩阵估计算法)
11. Fang, L.: Research on feature based 3D scene reconstruction techniques from image sequence. Huazhong University of science and technology (2007), (基于特征的图像序列三维场景重建技术研究)
12. Boissonnat, J.D.: Geometric structures for three-dimensional shape representation. *ACM Transactions on Graphics*. 3(4), 266–286 (1984)
13. Wang, Q., Wang, R.Q., Ba, H.J., Peng, Q.S.: A Fast Progressive Surface Reconstruction Algorithm for Unorganized Point. *Journal of Software*. 11(9), 1221–1227 (2000), (散乱数据点的增量快速曲面重建算法)
14. Hou, W.G., Ding, M.Y.: Method of Triangulating Spatial Point s Based on Manifold Stud. *J. Acta Electronica Sinica* 37, 2579–2583 (2009), (基于流形学习的三维空间数据网格剖分方法)
15. Li, L.D., Lu, D.T., Kong, X.Y., Wu, G.: Implicit Surfaces Based on Radial Basis Function Network. *Journal of Computer-aided Design & Computer Graphics* 18, 1142–1148 (2006), (径向基函数网络的隐式曲面方法)
16. Fang, L.C., Wang, G.Z.: Radial basis functions based surface reconstruction algorithm. *Journal of Zhejiang University (Engineering Science)* 44, 728–731 (2010), (基于径向基函数的曲面重建算法)
17. Tenenbaum, J.B., Silva, V.D., Langford, J.C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction 290(5500), 2319–2323 (2000)

Distribution and Rendering of Real-Time Scene Based on the Virtual Reality System

Chengfang Zhang¹ and Guoping Wang^{2,3}

¹Shenzhen Graduate School, Peking University, Shenzhen, China

²Graphics and Interaction Lab of EECSS, Peking University, Beijing, China

³Beijing Engineering Technology Research Center of Virtual Simulation and Visualization,
Beijing, China

{zhangcf, wgp}@pku.edu.cn

Abstract. There are differences between real and virtual scene information displayed on a virtual reality platform, not a true reflection of the reality of the real-time information. On the basis of virtual reality platform, how to collect, organize and publish the corresponding real-time virtual reality scene information is becoming a new problem. In this essay, we will explain how to input, distribute and display real scene information based on the virtual reality system, a distributed virtual reality system. The current real-time video information corresponded with the specified scene object in the virtual reality system is recorded by the client, released to the streamer server and then distributed via content distribution server. The live streaming information displayed on the virtual reality client increases the fidelity and real-time of the virtual reality. The system is timely published, scalable and is capable of supporting remote deployment and distribution. The goal is to provide a reliable and effective real scene information dissemination and presentation of the real scene platforms based on the virtual reality environment.

Keywords: Virtual Reality, Streaming, Virtual Scene and Distribution.

1 Introduction

VR (Virtual Reality) is a computer-simulated environment that can simulate physical presence in places in the real world or imagined worlds. Most current virtual reality system focuses on the construction and rendering of reality. However, due to the problems of modeling, rendering and historical versions, the system does not truly reflect the information of real-world. The real-time information based on the real scene displayed and rendered in a virtual reality platform will be very meaningful. Virtual reality system plays a role not only to simulate the real-world simulations, but also to release and render real-world information, thereby improving the usability of the system.

With the popular use of mobile devices, the application based on social relations such as micro-blog and micro-visual is becoming a new way to publish and broadcast information. The real-time media increases the speed of information-spreading. But there is a limitation for expressing the dimension and position of the scene in the

information. With the support of geography, topography and building models in the virtual reality platform, the realistic scenes are displayed more intuitively and spatially.

The main contents of this essay is to display and render live streaming as the real-time scene through the semi-automatic correction corresponding to the scene based on ViWo[1]. The video streaming released on the corresponding location in the virtual reality platform is sent to the streaming media server. Then the streaming is processed to support multi-platforms and will be sent to the content distribution server. The video streaming is projected on the ViWo client, enhancing the actual and real-time social information and real scenes information. Meanwhile, the data caching and load balancing of the distribution server is designed and implemented for providing a steady stream of small delay.

2 Related Work

In this section we provide an overview of the relevant literature on the blending rendering of virtual scene with real scene. And then we introduce the distribution methods of the live streaming as the real scene and the replacement policy of the content distribution server.

2.1 Combine Virtual Scene with Real Scene to Render

There are two main ways to combine virtual scenes and real scenes to render in the virtual reality system. Augmented Reality (AR) [2] is a method of registering true 3D virtual object into the world to display or output video through. Since a large amount of positioning data analysis and computation is required to ensure that the scene information from the computer-generated virtual objects can be accurately positioned in the real scene, this method is not suitable for real-time recording of real scene information for the mobile device client. Another way is to increase the description of the scene node in the virtual environment, including the text, images and video. Additional information corresponding to the scene objects is a supplement for the virtual scene rendering. Icon is added in the Google earth as entrance of real scene's display. Although this approach based on the position of the virtual reality system is an effective way to organize and display the real scene, it lacks the real-time information for the video offline uploaded. The Real-time video streaming can effectively make up the information presented in real time.

2.2 Distribution Methods of Live Streaming

The way of traditional live streaming transmission consists of IP multicast transmission[3], p2p (peer to peer) transmission [4] and end-to-end transmission. IP multicast transmission is a method of sending IP datagrams to a group of interested receivers in a single transmission, allowing nodes in the network to get data from the client which has received data. This technology can reduce the pressure of the server and save the

network bandwidth, but it is not advanced enough to support the different equipment, which is not widely used in the live streaming transmission. P2p transmission is a way to share resources between client nodes which emphasize the equivalence of node, and the advantages of low cost, high scalability, high service quality and high security. However, it will increase the time delay for client and is difficult to support to support different clients as well. End-to-end transmission is a method that allows each client to build connection with the server and request data from the server. There's an advantage of fast response and strong adaptability. But there is a big pressure on the server. The layered content delivery network [5-7] based on the end-to-end transmission can effectively solve the problem of the pressure on the server and network limitation. Edge delivery node deployed in the system as the server is directly requested by client. Request delay and stress to backbone is reduced because there is a cache of hot data in the edge server.

2.3 Replacement Policy

As cache size is finite for the edge content distribution server, a cache replacement policy is needed to manage cache content. If the data flow from the media server needs to be stored when the cache is full, the policy will determine which data is evicted to make room for the new data. The Replacement policy is generally classified as time-based policy, frequency-based policy, size-based policy, function-based policy and randomized policy [8]. Hit rate and complexity are the main factors to be considered. Latest recently used (LRU) as time-based policy and least frequent used (LFU) as frequency-based policy are often employed in the web cache replacement policy. Size-based policy is used when large file is less popular. Larger file will be removed to contain more files with small size if data needs to be replaced. Function-based policy performs best but has the highest complexity. Greedy dual size policy [9] is a function-based policy that the files with larger fetch costs will be stated in the cache longer. Randomized policy works the worst so that it is not used in practice.

3 The Overall Architecture Design

Different from the structure of traditional content distribution network, the data is mostly collected and distributed according to the region. The design of the system should meet the following characteristics: one is easy to deploy and extend on the existing server structure, no need to alter the architecture of the virtual reality system; the second is able to quickly respond to the large-scale users' requests by increasing the hardware; the third is to support multiple platforms of virtual reality, including the current virtual reality client and mobile virtual reality client; the last one is to provide a stable flow of live streaming video and reduce the delay. The system consists

of video source client, code server, streaming media server, data server, content distribution server, virtual scene server and ViWo client. The video source is divided into fixed position monitor video source and mobile virtual reality client. The coding server is responsible to code the raw data from the fixed position camera and transmit code data to media server. As for the mobile devices, the local coding is used because of the bandwidth limit. The coded data is sent to the media server which is responsible for packing the live stream, storing data flow and dispatching data to the edge server node. At the same time, all the streaming resource is registered on the virtual scene server through communication with the media server. The client obtains resource list from the virtual scene server and then requests streaming flow from the edge content delivery server. When the edge content delivery server is requested for data flow, resource is checked if existed in the local cache. If the data is existed in the local, data will be directly read and then sent to client. Otherwise, the edge content delivery server will request data from the media server. This way Request delay and stress to backbone is reduced for a large number of requests.

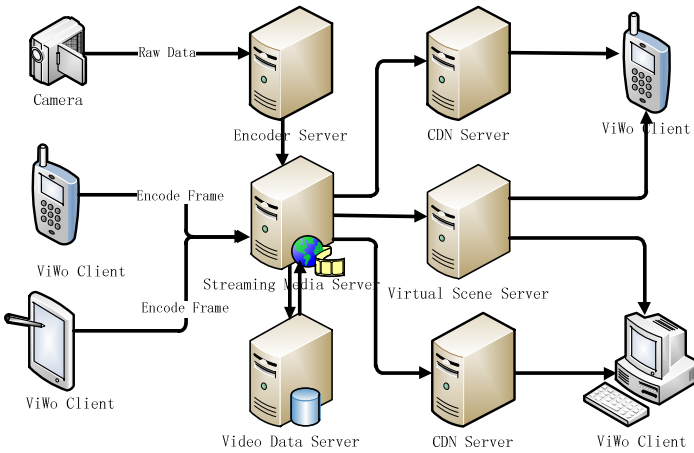


Fig. 1. The overall system design

3.1 Design for Streaming Media Server Module

As mentioned in the overall architecture above, the streaming media server in the system is responsible for data reception, stream Package, preservation and distribution. The video frame encoded can be received according to a uniform protocol and it can also increase the expansion sexual system access of the device. The coded data flow will be packed to flash video tag. It's important to indicate that the flash video tag is organized to a Group of Picture (GOP) structure that is started at key frame and maintains about a size of 40 frames. The index file that records GOP file location and timestamp is saved in the memory to response to client's request for flash video through the protocol of RTMP^[10]. Similar to the protocol RTMP, the encoded frame also needs to be packed to MPEG-TS slice, the file of m3u8 format that records the

path of slice file is transported through the protocol of HLS [11]. At the same time, encoded data is stored in the local data server. If the live streaming is finished, media data saved in the data server can also be requested by the client.

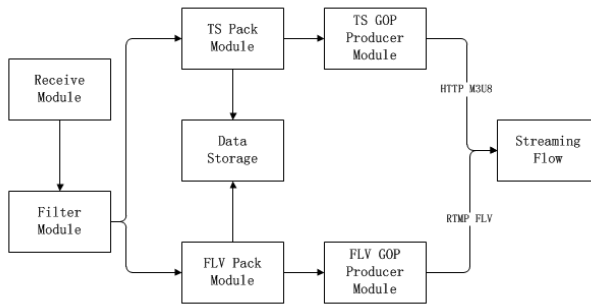


Fig. 2. Modules of media server

Table 1. Request method for RTMP and HLS protocol

URI	Explanation
rtmp://ip:port/chanel/datarate.flv/live	Flash live stream on specified channel and datarate
rtmp://ip:port/video/filename.flv&dr=datarate	Flash Video stream on specified datarate
http://ip:port/chanel/datarate.ts/live	Mpeg-tsLive stream on specified channel and datarate
http://ip:port/video/filename.m3u8&dr=datarate	Mpeg-ts video stream on specified datarate

3.2 Data Caching and Load Balancing for Distribution Server

When the edge distribution server is requested by client, the cache is checked if the data exists in the local cache. In the virtual reality system, if the request is hit on the local server, the data flow will be directly sent to client. Otherwise, the edge distribution server will turn to the central media server to request data. The requested data flow is saved in the local server so that the edge distribution server doesn't need to request data if next client requests the same data flow. In case there is no enough storage space for the cache of the data flow in the local server, the resource that is not recently used will be replaced. Replacement policy in the virtual reality system is more complicated than the traditional content distribution network because the data is continuously requested within the region. The visited frequency of Virtual scene, last request time of media streaming and times of recently visited are all key factors

for the replacement strategy. The best method is to replace the file video file that will not be visited in the future. We have calculated the weight of every media file in the cache server. The file with minimum weight will be replaced. F_v is the visited frequency of the virtual scene in the same zone with real scene. n is the number of visits of the media file. t_{visit} is the time of media file last visited. Based on information of the visit of the real scene and virtual scene, we have calculated the weight of the media file.

$$\omega = F_v * n * t_{visit} \quad (1)$$

The edge distribution server that has the smallest load will be selected and returned in the same subnet with the client when client requests data flow from server. Factors such as bandwidth, memory, I/O and CPU resources are taken into account for the content distribution server. However, bandwidth and memory are the two most important factors for the pressure from the video streaming distribution system. We have noticed that γ is the load factor to server, which is the sum of bandwidth factor α and memory factor β . B_{total} is the total amount of bandwidth of the server and b_i is the amount of bandwidth used by a client connection. M_{total} is the total amount of physical memory of the server and m_i is the amount of memory used by a client connection.

$$\gamma = \delta \cdot \alpha + (1 - \delta) \cdot \beta \quad (2)$$

$$\alpha = \frac{B_{use}}{B_{total}} = \frac{\sum_{i=1}^n c_i \times b_i}{B_{total}} \quad (3)$$

$$\beta = \frac{M_{use}}{M_{total}} = \frac{\sum_{i=1}^n c_i \times m_i}{M_{total}} \quad (4)$$

The content distribution server is dynamically allocated to achieve the load balance by controlling the memory factor and bandwidth factor. Because the server network environment, the operating system and other factors, the bandwidth normally take maximum factor 0.8, the memory factor is 0.6^[13]. One of the minimum load Servers in the region determined by IP address of client which is returned when client requests data flow. The server that γ is the minimum value is selected. More concurrent requests are carried and a steady stream of small delay has been caused by the strategy.

4 Collection and Rendering

The mobile device can identify the location of the virtual system by its GPS sensor when during the recording of a real scene. Because of the difference between the virtual scene and the real scene, the automatic calibration has the problem of high error

rate and more time consuming. We choose the method with which we manually adjust the angle of view on the real scene under the condition of the same position and the angle with the virtual scene. The virtual scene and the real scene are draw simultaneously at the same time during the process of filming. The parameter of alpha can be controlled during the process, and this makes it easy for user to correct the angle position between the video object and the virtual object. The real scene recorded at the fixed position is encoded by the encoding server. Then the video streaming is uploaded to the streaming media server. The virtual reality system will register the corresponding scene objects at a certain position. It means that there is a real scene monitoring at this position. Local coding is adopted for the mobile device because of the limitation of bandwidth. Thanks to the quality of the video signal, the resolution and client code flow both can be controlled. The acquisition module is not adopted to the adaptive network transmission encoding. The users can manually adjust the video coding rate according to the network transmission factor.

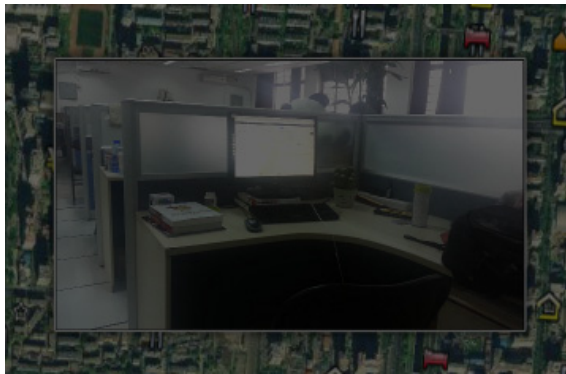


Fig. 3. Real scene indoor displayed on the mobile device



Fig. 4. Real scene outdoor rendered in the system

We have implemented the video rendering in the virtual reality platform including windows platform and android platform. Please note that not all the real video scene is easy to record on the alignment angle position. Two ways can be selected to render in the virtual reality system. As shown in figure 3, the real scene indoor is displayed in a dialog by clicking the camera icon. For the real scene that is aligned in the accurate angle position will be projected in the virtual reality system, which is more suitable for outdoor scenes. Figure 4 has showed the live video projected in the system. The scene object in the projection range shows the real situation of the current position. We can also see the object out of the projection range in the virtual reality system.

5 Experimental Evaluation

In this section we present a performance of hit rate experiment and load balancing test in the current distributed architecture. The Experimental environment of the system including the followings: a dell PowerEdgeR710 server (Intel Xeon 2.40GHz , 16GB), as a streaming media server; a data server (Intel Core i5 3.10GHz, 16GB); three distributed server (Intel Core 2 Duo 2.80GHz, 8GB); a test server (Intel Core 2 Duo 2.93GHz, 4GB); a Samsung Galaxy i9300 mobile client and the HTC Onex mobile client.

5.1 Hit Rate Experiment

The 40 GB disk space in the content disputation server is allocated to the cache video data for experiment. 400 requests with more than 200 different requests from 10 users are selected to make a simulation experiment. The size of the file is between 200MB and 1G. The Cache is full and the contents are the same for each experiment. Hit rate is counted with the function policy designed, LRU policy, LFU policy and size policy. As shown in figure 5, the weight function design has a better performance because its hit rate is higher than others.

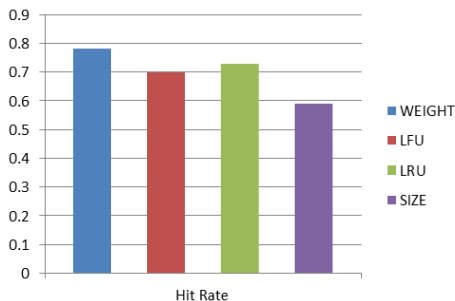


Fig. 5. Hit rate in four experiments

5.2 Load Balancing Experiment

The Distribution network bandwidth is set to be 100Mb/s to make up a test. The network bandwidth factor is set to 0.6. The memory factor is set to be 0.5. The distributed server is divided into two groups, which are in two subnets: subnet one consists of a distributed server, the second subnet consists of two distributed servers. The Samsung galaxy i9300 is used for the client to collect and transmit the video streaming. The frame rate is set to be 20. The resolution is set to be 320 * 240. The coding rate is set to be 300 kb/s. The WIFI network upload bandwidth is set to 1 MB/s.

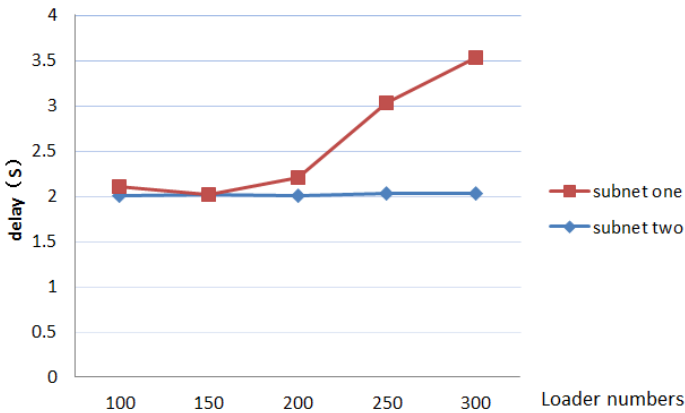


Fig. 6. The delay time at different pressure requests

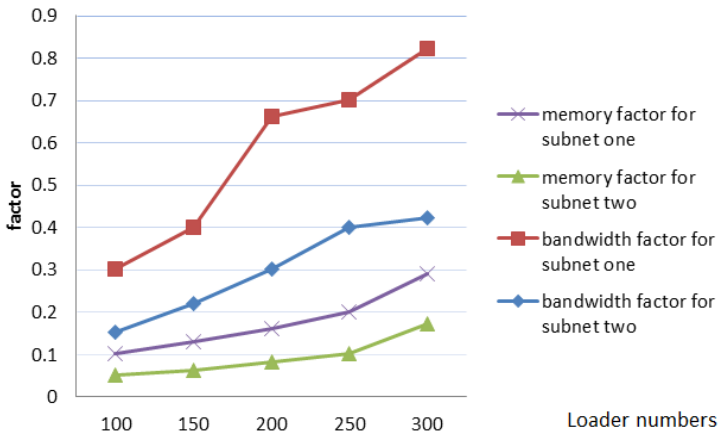


Fig. 7. The load factor at different pressure requests

The testing machine starts the video streaming at the rate of 300 Kb/s when real-time video is uploaded. The video is recorded 60 second. The delay time is counted by HTC Onex's request for the live streams when the pressure request number is 100, 150, 200, 250 or 300.

Analysis: As shown in figure 6, memory factor is below 0.5. Bandwidth factor is the key problem for the data request. When the mobile client is uploading video frames, the procedure of frame processing and saving will lead to a delay of 1s so that the delay is more than 2s at least. When the number of pressure test connections of the delay figure of subnet one is under 200, the delay time is about 2 second. When the pressure test connection number is more than 200, the bandwidth of the server has been filled so that the delay time is increased because there is only a distributed server in subnet one. However, there are enough bandwidths for second subnet so that the delay time is still 2s.

6 Conclusions

In this research we have presented a system that includes video recording, transmission, distribution and rendering based on the design and implementation of large-scale distributed virtual reality system. The real-time video scenes through mobile client are browsed or projected at the alignment angle position in the virtual reality. The data cache replacement policy and load balancing policy is designed for the content distribution server. In the real application environment, delay is increased compared to the video chat application because of the architecture of http server transmission as well as the delay of transcoding. Meanwhile, the live video scene is only displayed and projected in the virtual scene at the alignment position in the current system. It's better to use a good algorithm to blend virtual scene with real scenes. We believe that all these questions are worth a continued exploration in the future studies.

Reference

1. Wang, G.: ViWoSG: Ultra-large-scale Distributed Virtual Environment. SCIENCE CHINA Information Sciences 39(1), 96–106 (2009)
2. Caudell, T.: AR at boeing (EB/ OL) (1990), <http://www.ip0.tue.nl/homepages/mrauterb/presentations/HCI2history/tsld096.html>
3. Quinn, B., Ajmeroth, K.: IP multicast applications: Challenges and Solutions (EB/OL), <http://www.ietf.org/rfc/rfc3170.txt> (February 20, 2011)
4. Tran, D.A., Hua, K.A., et al.: A Peer-to-peer Architecture for Media Streaming. IEEE Journal on Selected Areas in Communications 22(1) (2004)
5. Vakali, A., et al.: Content Distribution Networks Status and Trends. IEEE Internet Computing, 68–74 (November 2003)
6. Mulerikkal, et al.: An Architecture for Distributed Content Delivery Network. IEEE (2007)

7. Lazar, I., Terrill, W.: Exploring Content Delivery Networking. *IT Professional* 3(4), 47–49 (2001)
8. Wong, K.-Y., Rao, A., Lanphier, R.: Web Cache Replacement Policies: A Pragmatic Approach, Macao Polytechnic Institute. *IEEE Network* (January 2006)
9. Cao, P., Irani, S.: Cost-Aware WWW Proxy Caching Algorithms. In: *Proc. USENIX Symp. on Internet Tech. and Sys.*, pp. 193–206 (December 1997)
10. Adobe System Inc. Real-Time Messaging Protocol Specification (2009)
11. Pantos, R., May, W.: HTTP Live Streaming. IETF Draft (June 2010)
12. Tian, X., Chen, S.: Proxy Cache Replacement Algorithms for Streaming Media Based on Smallest Cache Utility. *Journal of Computer Applications* 27(3), 733–736 (2007)
13. Yu, J., Liu, W., Wang, T.: Survey on Proxy Caching Technologies for Streaming Media. *Computer Science* 33(1), 18–21 (2006)

Probabilistic Model for Virtual Garment Modeling

Shan Zeng, Fan Zhou, Ruomei Wang, and Xiaonan Luo

National Engineering Research Center of Digital Life, State-Province Joint Laboratory of Digital Home Interactive Applications, School of Information Science Technology, Sun Yat-sen University, Guangzhou 510006, China
isswrm@mail.sysu.edu.cn

Abstract. Designing 3D garments is difficult, especially when the user lacks professional knowledge of garment design. Inspired by the assemble modeling, we facilitate 3D garment modeling by combining parts extracted from a database containing a large collection of garment component. A key challenge in assembly-based garment modeling is the identifying the relevant components that needs to be presented to the user. In this paper, we propose a virtual garment modeling method based on probabilistic model. We learn a probabilistic graphic model that encodes the semantic relationship among garment components from garment images. During the garment design process, the Bayesian graphic model is used to demonstrate the garment components that are semantically compatible with the existing model. And we also propose a new part stitching method for garment components. Our experiments indicates that the learned Bayesian graphic model increase the relevance of presented components and the part stitching method generates good results.

Keywords: Garment modeling, Bayesian Probabilistic Graphic Model, and Part Stitching.

1 Introduction

Modeling garments is essential for virtual fitting and can benefit other applications such as films and games. The modeling garment is a challenge task because the garment is complex in structure: it contains different kinds of components and varies of garment style. Berthouzoz [1] created detailed garment by seaming 2D patterns automatically. Bradley, Li, Zhou [2,3,4] model the garment from photos, it relaxes the professional requirement of garment design. However, their model couldn't model detailed garment that contains pocket, belt, and button, etc.

A successful 3D garment design system should be simple, intuitive to use, and provides multiple design options for different design intent. Assembly-based modeling provides a promising new approach to 3D garment modeling. If we have a database of garment components, we can model garments by assembling the components. Identification of relevant components to be presented to the user is a key problem in assembly-based modeling [5,6,7]. The advantage of assembly-based modeling is that users do not need to design 2D patterns.

In previous work, Chaudhuri, Kalogerakis [7,8] analyzed a large set of segmented 3D models and used a probabilistic graphic model to learn their semantic and geometry relations for the exploration of a smarter design. As this work relies on the existed 3D segmented models, we bypass the difficult work by learning from a large set of garment images.

In this paper, we present a probabilistic graphic model called Bayesian network [9,10] to automatic recommend garment components to the users during 3D garment modeling process. A probabilistic graphic model is well-suited to encoding the relationship between the random variables such as the garment style, sleeve style, collar style, pocket style, existence of belt and button, etc.

Our main contribution is that we propose a part assembly method for garment design, and we lean a probabilistic graphic model from images. The system automatically recommends garment components and we also propose an improved mean value coordinates [11] method for part stich.

We demonstrate the effectiveness of our model using the part assembly garment modeling tool that we have developed. Experiment shows that the probabilistic model produces more relevant recommendations than a static presentation of components could.

2 Related Work

Part assembly modeling remains a popular research topic, but we built our work on some of the ideas and algorithms. In this section, we present a few representative papers relevant to our work.

2.1 Part Assembly Modeling

As the model collections grow, the assembly-based modeling provides a quick way to create new models by reusing the existing models. The modeling-by-example system relies on shaped-based or text-based search to retrieve component parts [5]. [12,13]propose various sketch-based user interfaces: In these methods, the user must be very clear about the specific component. This is not appropriate for 3D garment modeling when the user has no design expertise. Kreavoy [6] described a method in which the user can interchange parts between a set of compatible shapes. This method requires the shapes that share the same number of components. Chaudhuri [7] proposed a data-driven technique that can recommend components to augment a given shape. This method just considers the geometry feature but doesn't take into account the semantics of components.

2.2 Probabilistic Framework

Fisher [14] describes a probabilistic model for 3D scene modeling. The user drew a boundary box, the model extracts a model from database using the context

information. Chaudhuri and Kalogerakis [7,8] propose the probabilistic models for automatically synthesizing 3D shapes using training data. These models reflect the probabilistic graphic models that encoding semantic and geometry relationships among shapes through segmented 3D labeled models. However, the number of 3D garment models is limited, and it lacks diversity. The learned model may easily be over-fitting. We train our probabilistic model on large database of garment image, which solved the over-fitting problem.

2.3 Knowledge Acquisition from Clothing Image

The fashion coordinates system [15] uses the probabilistic topic models to learn information about coordinates from visual features in each fashion item region(e.g. tops, bottoms). Our work focus on the learning of the relationship among garment components.(e.g. collar, pocket, sleeve, and etc.) Liu, Yu[16,17] also study on outfit combination; and in our work , we consider the combination of garment components.

3 Overview

Our approach comprises an offline preprocessing stage in which a probabilistic graph model is trained on a garment image database and an online interactive stage in which the garment component parts are recommended and assembled. The overview of our work is illustrated in Fig.1.

Offline Preprocessing. In the offline stage, the outline of processing pipeline is as the followings:

- (a) Image collection and attributes annotation. Collecting garment images from online shopping websites. And manually annotated the garment components attributes, such as garment style, collar style, sleeve style and pocket style, etc.
- (b) Training . Learning the Bayesian graphical model which encodes the relationship of garment attributes.
- (c) 3D models cluster and annotation. The garment components are clustered by geometry features, and then be labeled with style attributes. Thus we can use the probabilistic graphic model learned in the training stage for components recommendation.

Runtime Modeling. In the garment modeling stage, when the user choose a garment body or the user select a garment style, the system update the components rank list and the component are stitched to the existing garment using our part stich method.

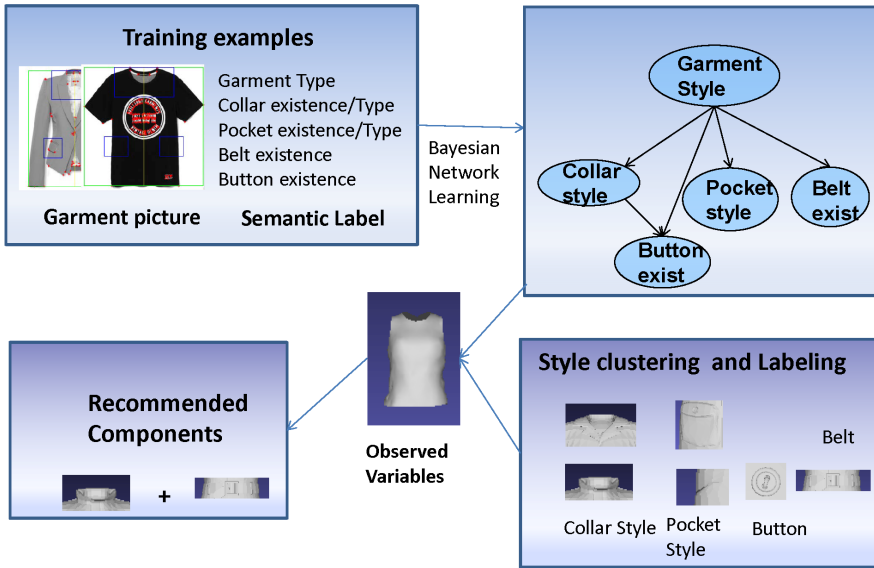


Fig. 1 system overview

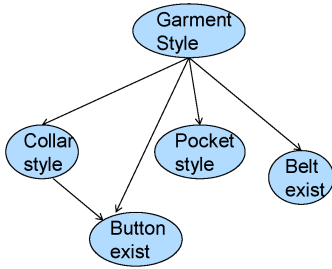
4 Bayesian Networks for Clothing Components Relationships

Clothing Image Dataset and Attributes Annotation. The garment images are crawled from the online shopping websites by using the keywords such as “T-shirt”, “coat”, and “suit”, etc. The images are classified by the keywords of style. The garment style is affected by the garment components attributes, for example, a T-shirt always has a short sleeve and has no button, a suit always has a flat collar and has buttons but do not have belt, etc.

We manually labeled the garment components. We define the garment component parts attributes and labeled about 500 clothing images for Bayesian network training. The attributes of collar, pocket, button, belt of each image are labeled.

Our system trains the Bayesian network for garment components for five garment styles: T-shirt, shirt, skirt, coat and suit. Figure 2 shows a part of the Bayesian network for garment components. The nodes of garment components have different attributes we have defined.

The node of Bayesian network represents different garment component parts and each node state represents the style attribute of garment component. For example, the node collar style has attributes stand-collar, fold-collar and flat-collar. And each component node has a state, but none of which indicates that the garment does not contain that component.



notation	domain	interpretation
G	$G \in Z^+$	garment style
C	$C \in \{0\} \cup Z^+$	collar style, 0 means no collar exist
P	$P \in \{0\} \cup Z^+$	Pocket style, 0 means no pocket exist
B	$B \in \{0,1\}$	0 means no button exist, 1 mean button exist
T	$T \in \{0,1\}$	0 means no belt exist, 1 mean belt exist

Fig. 2. Representing distributions of clothing components with a Bayesian network. Top: a Bayesian network for clothing components, trained by the labeled clothing image database. Bottom: a table showing the random valuables node and descriptions.

4.1 Probabilistic Graphic Model

Our probabilistic graphic model encodes the joint distribution on garment style and components. The purpose of the model is to recommend garment components compatible to the existing models. The hierarchical graphic model is showed in Fig.2. It contains G .as the root represents the garment style, and contains random valuables such as collar, sleeve, garment body, pocket, belt and button. which represent C, S, M, P, B, T respectively. $C, S, M, P \in \{0\} \cup Z^+$, 0 represents none, and nonzero represents the style attributes of the component. For $B, T \in \{0,1\}$ 0 means none and 1 means exist. As the belt and button style is not as complex as others, we reduce the model complexity just by considering their existence. The graphic model may contain lateral edges between the nodes representing the strong dependency between components, which can be learned from the structure learning.

The conditional distribution of discrete random variable of C, S, M, P, B, T can be represented as conditional distribution table(CPT). Considering a discrete variable Y with a single parent discrete variable X , each assignment y to Y and x to X the CPT at Y contains the entry:

$$P(Y = y | X = x) = q_{y|x} \tag{1}$$

The values $Q = \{q_{y|x}\}$ is the parameters of the CPT. For the root node G , the parameters comprises $P(G = g) = q_g$.

We define the joint distribution as $P(U)$, when a component is selected, the node is set as observed variable and the unobserved variables are query variables. For example, when the stand-collar is selected, C became observed and is assigned to 1, or if the garment style is set as T-shirt, the node G became observed.

During the inference process, when given the observed variables, we compute the probability of query variables. Assuming given O , which contains the subset of

$\{C, S, M, P, B, T\}$, $U_q \in U$, q is the query variable, and we define the score as the marginal probability:

$$score(U_q) = P(U_q = q | O) \quad (2)$$

Learning

The input of the offline learning process is a vector set D representing the labeled clothing image dataset. $D = \{C_i, S_i, M_i, P_i, B_i, T_i\}$, $i = \{1, 2, \dots, K\}$. We have K training images in total. The output of the learning process is a directed graph and the conditional probabilistic table(CPT) for each node.

The best structure G is the one that has highest probability when given training data D (Kollar and Friedman 2009). By Bayes' rule, the probability is as follows:

$$P(G | D) = \frac{P(D | G)P(G)}{P(D)} \quad (3)$$

Cooper [18] assumes a uniform prior distribution $p(G)$ over all possible structures. Thus maximizing $P(G | D)$ is equal to maximizing $P(D | G)$. We define θ as the prior distribution over structure G , then the marginal likelihood can be expressed as:

$$P(D | G) = \int_{\theta} P(D | G, \theta)P(\theta | G)d\theta \quad (4)$$

We adopt the K2 algorithm to the structure learning The K2 algorithm[18] is a greedy search algorithm that works in the following ways. Initially each node has no parents. It then incrementally adds the parent whose addition increases the score of the resulting structure most. When the addition of no single parent can increase the score, it stops adding parents to the node. After performing the K2 algorithm, the structure and the CPT are learned.

4.2 Clothing Components Recommendation

3D Model Clustering: the object of our system is applying the implicit relationships among the garment components learned from images to 3D detailed garment modeling. Firstly, we need to establish the relationship between the images and 3D models. During the training stage, we labeled the style attributes of each component, for example, the collar style is labeled as stand-collar, fold-collar and flat-collar. Thus, we also labeled the style attributes for each 3D garment component. To annotate the 3D garment component models, we use K-Means classifier to cluster the models. For 3D model feature extraction, we adopt the visual based method [19], and in his method, each model is rotated twelve times and be projected in 10 direction. Although it has high recall and precision, it costs much more time. In our system, we simplify this method by aligning the model first, we use PCA to find the three main component axis and get the project view from the three direction; And then we extract the shape moment feature for clustering.

Garment Parts Recommendation: The goal of our system is to facilitate the garment modeling process, and garment parts are automatically presented to the user. In our system, there are two ways to make recommendation. When the user select an item from the garment part database, the system recomputed the probability of each garment part style and present them to the user, or when the user specify the garment style, for example, the user specify a suit, then flat-collar, inserted-pocket, button may be recommended.

The recommendation decision is made according to the score computed by Bayesian inference according equation(2). O is the subset of $\{C, S, M, P, B, T\}$, $U_q \in U$, q is the garment part need to be assembled. For example, when the garment body is observed, M is assigned to H-style, if the query part is collar, then we have:

$$score(C_x) = P(C = x | M = H - style) \quad (5)$$

The collar style with the maximum score is recommended, namely:

$$\arg \max_x \{score(C_x)\} \quad (6)$$

5 Part Assembly

The user can select a part p from the recommended list, and stich it to the existing garment body. Our part stitching method can be handled with both closed loop boundary mesh such as pocket, sleeve and unclosed none-loop boundary such as collar. As for button, which has no boundaries, it can also work well.

Part Stitching. The user specifies the corresponding contact pair vertices and then the garment part can be automatically snapped to the garment body. We propose a cage-free method based on mean value coordinates to deform the source part. For both closed loop and non-closed loop parts, we use the user specified boundary vertex in the source mesh as the boundary vertex of the source cage. To preserve the normal of the source surface, we add a vertex in the direction of the average normal of the boundary vertex and another in the opposite direction with length of the source part's radius R .

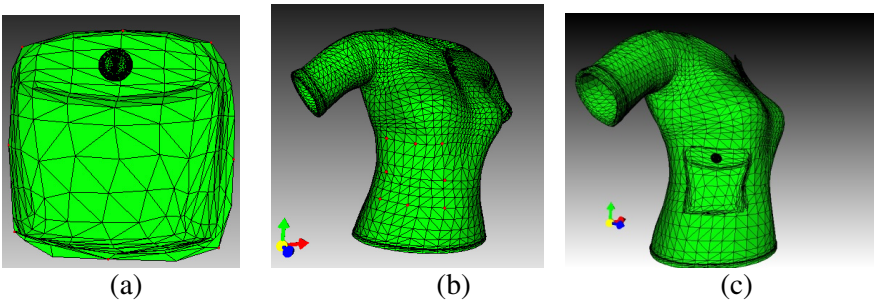


Fig. 3. (a) The pocket is the source part (b)The garment body is the target mesh to be assembled. (c)The result garment.

We use these two vertices and the selected boundary vertices construct the control cage of the source part.

Size Aware Deformation. Assume that the source part is S and the target part is T , we automatically fit the source S to the target T by computing the ratio k of source's boundary length and target's boundary length. Then we construct a new cage using the selected vertex in the target mesh, and add two vertices with length of kR in the same way as source cage dose. We deform the source part using the new control cage and interpret with the mean value coordinate computed by the old cage. The source mesh is stitched to the target mesh finally. See Fig.3.

6 Result and Discussion

We train the probabilistic graphic model on a cloth image dataset crawled from online shop. The cloth image contains five garment style, T-shirt, shirts, skirt, coat and suit. The learned structure is showed in Fig.2. The learned conditional probability of component node is showed in Table 1. Fig.4 shows that the learned Bayesian graphic model increase the relevance of presented components than static ordering.

Some of the generated garment model is showed in Fig.5. As we established a one to one mapping of the boundary vertex, garment components with border like skirt, collar and sleeve are merged into garment body seamlessly. Components are adaptively scaled to fit the garment body without distortion. Our cage-deformation method adds a control vertex in the normal direction; it can preserve well the surface detail feature of garment components well. Our deformation method adopts mean value coordinate to interpolate the new position of vertex, so it can also be applied to assemble components with complex topology like button and belt.

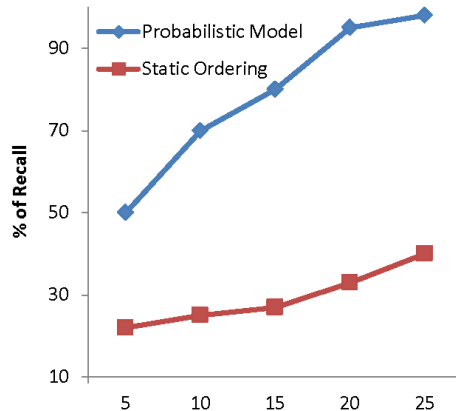
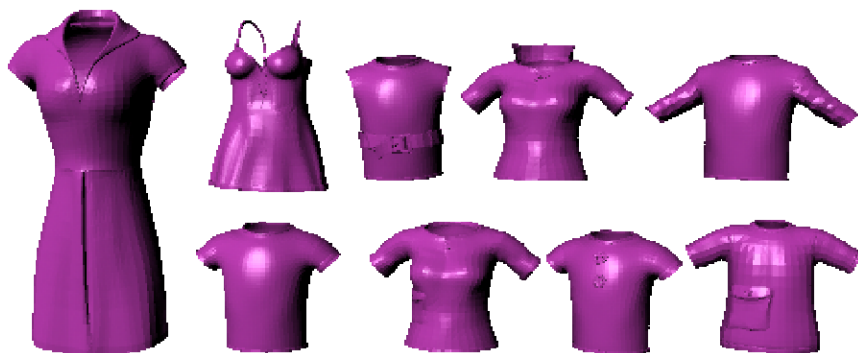


Fig. 4. The figure shows the recall of components, the probabilistic model presented more relevant components than the static ordering

Table 1. Some Results of the conditional probability table of component node

Garment Style	Collar Style				Pocket Style				Belt	
	No	Stand	Fold	Flat	No	Patch	Side	Insert	Exist	Not Exist
T-shirt	0.7	0.0	0.3	0.0	0.7	0.3	0.0	0.0	1.0	0.0
Shirt	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0
Skirt	0.9	0.0	0.1	0.0	0.9	0.1	0.0	0.0	0.6	0.4
Coat	0.2	0.1	0.7	0.0	0.1	0.4	0.1	0.4	0.9	0.1
Suit	0.0	0.0	0.0	1.0	0.0	0.6	0.4	0.0	1.0	0.0


Fig. 5. Some generated garment results using our method

7 Conclusion

We present a probabilistic virtual garment modeling method which facilitated the 3D detailed garment modeling. Our system can automatically present the garment components to user during the modeling process. The system can infer the implicit garment style and suggest related garment component part. We also propose a cage-free method based on a mean value coordinate for part stitching. This method is rotation and scale in-relevant. It reduces the part alignment process and can preserve the smoothness of the part surface. It is fast and effectively.

Currently, we have learned the Bayesian network which encoding the relationship among garment components from pictures, and we just take into account the garment component style attributes, but it is flexible to add additional attributes such as texture, material and shape, etc.

In the future, we plan to take into account the spatial relationship of garment components which may reduce the labor of specifying corresponding vertices and make the garment modeling process faster.

Acknowledgments. We would like to thank the anonymous reviewers for their valuable comments. We were supported by the National Natural Science Foundation of China (61379112, 61272192).

References

1. Berthouzoz, F., Garg, A., Kaufman, D.M., Grinspun, E., Agrawala, M.: Parsing sewing patterns into 3D garments. *ACM Transactions on Graphics (TOG)* 32(4), 85 (2013)
2. Bradley, D., Popa, T., Sheffer, A., Heidrich, W., Boubekeur, T.: Markerless garment capture. *ACM Transactions on Graphics TOG* (2008)
3. Li, W.-L., Lu, G.-D., Geng, Y.-L., Wang, J.: 3D Fashion Fast Modeling from Photographs. In: 2009 WRI World Congress on Computer Science and Information Engineering. IEEE (2009)
4. Zhou, B., Chen, X., Fu, Q., Guo, K., Tan, P.: Garment Modeling from a Single Image. *Computer Graphics Forum* (2013)
5. Funkhouser, T., Kazhdan, M., Shilane, P., Min, P., Kiefer, W., Tal, A., Rusinkiewicz, S., Dobkin, D.: Modeling by example. *ACM Trans. Graph.* 23(3), 652–663 (2004)
6. Krevoy, V., Julius, D., Sheffer, A.: Model composition from interchangeable components. In: 15th Pacific Conference on Computer Graphics and Applications, PG 2007. IEEE (2007)
7. Chaudhuri, S., Kalogerakis, E., Guibas, L., Koltun, V.: Probabilistic reasoning for assembly-based 3D modeling. *ACM Transactions on Graphics TOG* (2011)
8. Kalogerakis, E., Chaudhuri, S., Koller, D., Koltun, V.: A probabilistic model for component-based shape synthesis. *ACM Transactions on Graphics (TOG)* 31(4), 55 (2012)
9. Pearl, J.: Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann (1988)
10. Kollar, D., Friedman, N.: Probabilistic graphical models: Principles and techniques. The MIT Press (2009)
11. Floater, M.S.: Mean value coordinates. *Computer Aided Geometric Design* 20(1), 19–27 (2003)
12. Lee, J., Funkhouser, T.: Sketch-based search and composition of 3D models. In: Proceedings of the Fifth Eurographics Conference on Sketch-Based Interfaces and Modeling. Eurographics Association (2008)
13. Fisher, M., Savva, M., Hanrahan, P.: Characterizing structural relationships in scenes using graph kernels. *ACM Transactions on Graphics TOG* (2011)
14. Fisher, M., Hanrahan, P.: Context-based search for 3D models. *ACM Transactions on Graphics TOG* (2010)
15. Iwata, T., Watanabe, S., Sawada, H.: Fashion coordinates recommender system using photographs from fashion magazines. In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, vol. 3. AAAI Press (2011)
16. Liu, S., Feng, J., Song, Z., Zhang, T., Lu, H., Xu, C., Yan, S.: Hi, magic closet, tell me what to wear! In: Proceedings of the 20th ACM International Conference on Multimedia. ACM (2012)
17. Yu, L.-F., Yeung, S.-K., Terzopoulos, D., Chan, T.F.: DressUp!: outfit synthesis through automatic optimization. *ACM Transactions on Graphics (TOG)* 31(6), 134 (2012)
18. Cooper, G.F., Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data. *Machine learning* 9(4), 309–347 (1992)
19. Chen, D.Y., Tian, X.P., Shen, Y.T., Ouhyoung, M.: On visual similarity based 3D model retrieval. *Computer Graphics Forum* (2003)

Dense 3D Reconstruction and Tracking of Dynamic Surface

Jinlong Shi, Suqin Bai, Qiang Qian, Linbin Pang, and Zhi Wang

School of Compute Science and Engineering, Jiangsu University of Science and Technology,
Zhenjiang, China

Abstract. This essay addresses the problem of dense 3D reconstruction and tracking of dynamic surface from calibrated stereo image sequences. The primary contribution of this research topic is that a novel framework of 3D reconstruction and tracking of dynamic surface is proposed, where a surface is divided into several blocks and block matching in stereo and temporal images is used instead of matching the whole surface, when all the block correspondences are obtained, a special bilinear interpolation is applied to precisely reconstruct and track the integral surface. Performance is evaluated on challenging ground-truth data generated by 3D max, and then different surface materials, such as fish surface, paper and cloth are used to test the actual effect. The research results demonstrate that this framework is an effective and robust method for dynamic surface reconstruction and tracking.

Keywords: Dense, 3D Reconstruction, Tracking and Dynamic Surface.

1 Introduction

Dense 3D reconstruction and tracking of dynamic surface provides more dynamic information than reconstruction of static surface; this makes the former more useful in many applications such as animation, motion capture and medical analysis. Animations require the real appearance of real-world objects from multi-view video for simulating the real scenes [1, 4, 22]. When researchers capture and analyze motion of objects, it is very important to collect accurate data of the distance and orientation of motion [19]. And in some medical fields, it seems the most practical solution is to use vision-based techniques for tracking heart motion [20, 8, 11], or establishing virtual environment of surgeon [14, 13].

However, dense 3D reconstruction and tracking of dynamic surfaces have not reach the satisfying level for the current stereo vision methods, and the primary problem, namely how to precisely perform matching between stereo and temporal images, is still tough in 3D reconstruction and tracking.

Researchers have proposed a variety of methods for reconstruction of 3D surfaces. The two types of commonly used methods are marker-based methods [17, 6, 21] and marker-less methods [18, 12, 7, 5, 3]. Marker-based method uses reflective markers or special regular textural markers attached to the surface, and track these markers in calibrated images. But the accuracy is limited by the number of the markers and their

weight. Marker-less methods is a very attractive non-invasive approach since it is not restricted to motion information associated with markers. However, for both marker-based and marker-less techniques, most of them provide surface reconstruction without spatio-temporal coherence which is usually important in many applications.

Recently, some researchers focus on optical flow method to reconstruct and track the deformable surface. For example, [9] provided an optical-flow based approach for deformable surface tracking using a mesh based deformation model together with smoothing constraints that force the mesh to shrink instead of fold in presence of self-occlusion, and [10] used optical flow to estimate scene flow from a calibrated stereo image sequence. Despite of their success in obtaining dense data of dynamic surface, they are easy to be influenced by illumination, and the process is usually involved in intensive computation.

Some other approaches [15, 16] tracked the deformable surface represented by a triangulated mesh where the problem is formulated as Second Order Cone Programming (SOCP). Though these methods can obtain good results, some prior knowledge of the possible deformations, such as the pose in the first or each frame of the deformable surface, should be required.

In this essay, we present a novel spatio-temporal framework of dense 3D reconstruction and tracking of dynamic surface from a calibrated stereo image sequence using block matching, by extending the Lucas-Kanade method [2] into the spatio-temporal domain. In our approach, the surface in one image can be divided into several blocks, and the counterparts of each block can be found by optimizing an energy function in the stereo and temporal images. Through this way the block correspondences of whole image can be obtained, so the surface can be reconstructed and tracked densely. However, due to the different parameters in different neighboring blocks, gaps may exist when we put together different reconstructed blocks. To solve this problem and get a smooth and dense surface, a bilinear interpolation method is adopted.

2 The Proposed Method

2.1 Description of the Notations

There are a lot of notations (listed in the following table) used in this research paper.

2.2 Divide One Surface into Several Blocks

We consider four images in a stereo image sequence, two images in the left image sequence and two in the right image sequence. The previous left image is called template image because it is the benchmark image for finding the correspondences in the four images, and the other three images will project back to the template image. One surface in template image can be divided into several blocks for every block

Table 1. Description of notation

Notation	Description
I_{pl}	The previous left image, we also call it template image in this paper.
I_{pr}	The previous right image.
I_{nl}	The next left image.
I_{nr}	The next right image.
B_{pl}	The block in I_{pl} , we also call it template block in this paper.
B_{pr}	The block in I_{pr} .
B_{nl}	The block in I_{nl} .
B_{nr}	The block in I_{nr} .
X	$(x, y)^T$, The coordinates in B_{pl} .
a	$(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6)^T$, is the affine parameter between blocks in the previous left image B_{pl} and the one in the previous right image B_{pr} .
$\Delta\alpha$	The increment of the affine parameter a .
d_L	The shift between the left two blocks.
Δd_L	The increment of d_L .
d_R	The shift between the right two blocks.
Δd_R	The increment of d_R .
$A(X; a)$	$\begin{pmatrix} (1+\alpha_1) & \alpha_3 & \alpha_5 \\ \alpha_2 & (1+\alpha_4) & \alpha_6 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$, denote the affine transformation.
$S(X; d)$	The shift function between the previous and next images.

in the template image; its corresponding blocks in the other three images can be located by the proposed method. Fig.1 illustrates the block matching process. After all the block correspondences are obtained in the surface, the whole surface can be reconstructed and tracked. So, the main problem is how to find one block correspondence.

2.3 Match One Block in Spatio-Temporal Images

Our matching process is based on the assumption that for small blocks in the images, only the affine transformation exists between the blocks of left image and the corresponding right one, and only the shift exists between the successive blocks. Under this hypothesis, the problems come down to find the affine parameters between left and right block, and shift parameters between previous and next block.

The Energy Function of Match One Block An energy function is designed in terms of the sum of squared error among the four blocks of the four images to find the block correspondence. This energy function includes three components, as shown in Eq.(1), namely the sum of squared error between B_{pl} and B_{pr} , the sum of squared error between B_{pl} and B_{nl} and the sum of squared error between B_{pr} and B_{nr} .

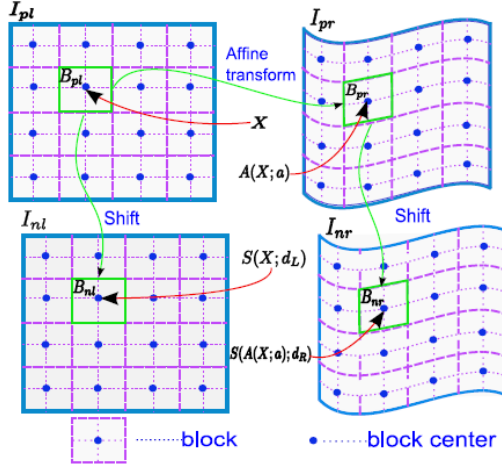


Fig. 1. Block matching in stereo and temporal images

$$\begin{aligned}
 E = & \sum_X [B_{pr}(A(X; \alpha + \Delta\alpha)) - B_{pl}(X)]^2 + \sum_X [B_{nl}(S(X; d_L + \Delta d_L)) - B_{pl}(X)]^2 \\
 & + \sum_X [B_{nr}(S(A(X; \alpha + \Delta\alpha); d_R + \Delta d_R)) - B_{pr}(A(X; \alpha + \Delta\alpha))]^2
 \end{aligned} \quad (1)$$

We assume that the current estimates of α , d_L , d_R is known, and all their initial values are set to zero, then iteratively solves for increments to the parameters $\Delta\alpha$, Δd_L , Δd_R , then update the estimates to α , d_L , d_R until the estimates of the parameters converge. The minimization of the expression in Eq.(1) is performed with respect to α , d_L , d_R , $\Delta\alpha$, Δd_L , Δd_R and the sum is performed over all of the pixels X of B_{pl} . The non-linear expression in Eq.(1) is linearized by performing a first order Taylor expansion leading to Eq.(2), Eq.(3), Eq.(4).

$$\sum_X [B_{pr}(A(X; \alpha + \Delta\alpha)) - B_{pl}(X)]^2 = \sum_X \left[B_{pr}(A(X; \alpha)) + \nabla B_{pr} \cdot \left(\frac{\partial A}{\partial \alpha} \right) \cdot \Delta\alpha - B_{pl}(X) \right]^2 \quad (2)$$

$$\sum_X [B_{nl}(S(X; d_L + \Delta d_L)) - B_{pl}(X)]^2 = \sum_X \left[B_{nl}(S(X; d_L)) + \nabla B_{nl} \cdot \left(\frac{\partial S}{\partial d_L} \right) \cdot \Delta d_L - B_{pl}(X) \right]^2 \quad (3)$$

$$\begin{aligned}
 & \sum_X [B_{nr}(S(A(X; \alpha + \Delta\alpha); d_R + \Delta d_R)) - B_{pr}(A(X; \alpha + \Delta\alpha))]^2 \\
 = & \sum_X \left[\left(B_{nr}(S(A(X; \alpha + \Delta\alpha); d_R)) + \nabla B_{nr} \cdot \left(\frac{\partial S}{\partial d_R} \right) \cdot \Delta d_R \right) - \left(B_{pr}(A(X; \alpha)) + \nabla B_{pr} \cdot \left(\frac{\partial A}{\partial \alpha} \right) \cdot \Delta\alpha \right) \right]^2 \quad (4)
 \end{aligned}$$

In the previous expressions, $\nabla B_{pr} = \left(\frac{\partial B_{pr}}{\partial x}, \frac{\partial B_{pr}}{\partial y} \right)$ is the gradient of block B_{pr} evaluated at $\mathcal{A}(X; \alpha)$, and ∇B_{pr} is first computed in the coordinate frame of B_{pr} and then warped back onto the coordinate frame of B_{pl} using the current estimate of the warp $\mathcal{A}(X; \alpha)$.

$\nabla B_{nl} = \left(\frac{\partial B_{nl}}{\partial x}, \frac{\partial B_{nl}}{\partial y} \right)$ is the gradient of block B_{nl} evaluated at $\mathcal{S}(X; d_L)$, and ∇B_{nl} is first computed in the coordinate frame of B_{nl} and then shifted back onto the coordinate frame of B_{pl} using the current estimate of the shift $\mathcal{S}(X; d_L)$. $\nabla B_{nr} = \left(\frac{\partial B_{nr}}{\partial x}, \frac{\partial B_{nr}}{\partial y} \right)$ is the gradient of block B_{nr} evaluated at $\mathcal{S}(A(X; \alpha + \Delta\alpha); d_R)$, and ∇B_{nr} is first computed in the coordinate frame of B_{nr} and then shifted back onto the coordinate frame of B_{pr} using the current estimate of the shift d_R and then warped back onto the coordinate frame of B_{pl} using the current estimate of $\mathcal{A}(X; \alpha + \Delta\alpha)$.

The term $\frac{\partial A}{\partial \alpha}$ is the Jacobian of the affine, which is defined as $(A_x(X; \alpha), A_y(X; \alpha))^T$.

After the first order Taylor expansion of Eq.(1) is performed. The partial derivative of Eq.(1) with respect to $\Delta\alpha$, Δd_R and Δd_L can be achieved respectively. $\Delta\alpha$ can be solved by setting the partial derivative of Eq.(1) with respect to $\Delta\alpha$ to zero.

$$\Delta\alpha = H^{-1} \cdot \sum_X \left[\begin{array}{l} \left[\nabla B_{pr} \cdot \left(\frac{\partial A}{\partial \alpha} \right)^T \right] \cdot [B_{pl}(X) - B_{pr}(A(X; \alpha))] \\ + \left[\nabla B_{nr} \cdot \left(\frac{\partial S}{\partial d_R} \right) \cdot \left(\frac{\partial A}{\partial \alpha} \right) - \nabla B_{pr} \cdot \left(\frac{\partial A}{\partial \alpha} \right)^T \right] \\ \cdot \left[B_{pr}(A(X; \alpha)) - B_{nr}(S(A(X; \alpha); d_R)) - \nabla B_{nr} \cdot \left(\frac{\partial S}{\partial d_R} \right) \cdot \Delta d_R \right] \end{array} \right] \quad (5)$$

Where H is a 6×6 matrix:

$$H = \sum_X \left[\begin{array}{l} \left[\nabla B_{pr} \cdot \left(\frac{\partial A}{\partial \alpha} \right)^T \right] \left[\nabla B_{pr} \cdot \left(\frac{\partial A}{\partial \alpha} \right) \right] \\ + \left[\nabla B_{nr} \cdot \left(\frac{\partial S}{\partial d_R} \right) \cdot \left(\frac{\partial A}{\partial \alpha} \right) - \nabla B_{pr} \cdot \left(\frac{\partial A}{\partial \alpha} \right)^T \right]^T \\ \cdot \left[\nabla B_{nr} \cdot \left(\frac{\partial S}{\partial d_R} \right) \cdot \left(\frac{\partial A}{\partial \alpha} \right) - \nabla B_{pr} \cdot \left(\frac{\partial A}{\partial \alpha} \right) \right] \end{array} \right] \quad (6)$$

Similarly, Δd_R and Δd_L can be obtained, as shown in Eq.(7) and Eq.(9):

$$\Delta d_R = H_R^{-1} \cdot \sum_X \left[\nabla B_{nr} \cdot \left(\frac{\partial S}{\partial d_R} \right)^T \right] \cdot \left[B_{pr}(A(X; \alpha)) - B_{nr}(S(A(X; \alpha); d_R)) - \left(\nabla B_{nr} \cdot \left(\frac{\partial S}{\partial d_R} \right) \cdot \left(\frac{\partial A}{\partial \alpha} \right) - \nabla B_{pr} \cdot \left(\frac{\partial A}{\partial \alpha} \right) \right) \cdot \Delta\alpha \right] \quad (7)$$

Where H_R is the 2×2 Hessian matrix:

$$H_R = \sum_x \left[\nabla B_{nr} \cdot \left(\frac{\partial S}{\partial d_R} \right) \right]^T \cdot \left[\nabla B_{nr} \cdot \left(\frac{\partial S}{\partial d_R} \right) \right] \tag{8}$$

$$\Delta d_L = H_L^{-1} \cdot \sum_x \left[\nabla B_{nl} \cdot \left(\frac{\partial S}{\partial d_L} \right) \right]^T [B_{pl}(X) - B_{nl}(A(X; d_L))] \tag{9}$$

And H_L is the 2×2 Hessian matrix:

$$H_L = \sum_x \left[\nabla B_{nl} \cdot \left(\frac{\partial S}{\partial d_L} \right) \right]^T \cdot \left[\nabla B_{nl} \cdot \left(\frac{\partial S}{\partial d_L} \right) \right] \tag{10}$$

2.4 Reconstruct and Track the Whole Surface

After all correspondences of blocks are found in four images, reconstruction and tracking of block surfaces can be independently obtained through the parameters of each block correspondence, then the reconstructed block surfaces can be integrated into one integral surface. However, because the parameters may be different between different neighboring blocks, gaps may exist between the reconstructed surfaces of neighboring blocks, which leads to the reconstructed unsmoothed surface. To solve this problem, a bilinear interpolation method is adopted, as Fig.2 shows.

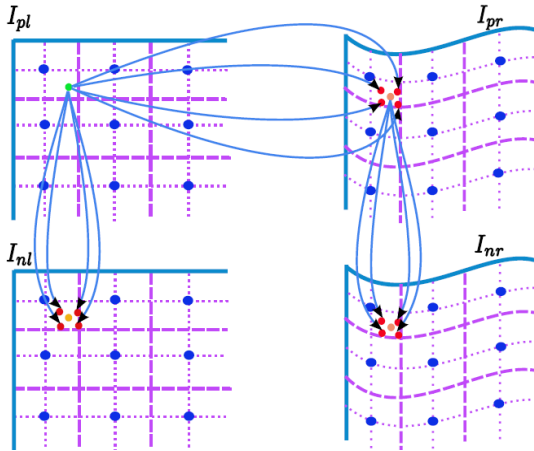


Fig. 2. Interpolation method of reconstruction and tracking

Bilinear Interpolation. Figure2 illustrates the bilinear interpolation method for smoothing reconstructed surface. For every four neighboring blocks in the template image I_{pl} , we consider the rectangle whose four vertices are composed of the centers of the four neighboring blocks, in which one point should be influenced by the parameters of the four neighboring blocks when transformation is carried out. As a result, four corresponding points in image I_{pr} can be generated by the affine parameters of the four neighboring blocks for every point in the rectangle. However, due to the different values of the parameters of the four neighboring blocks, the corresponding points in the image I_{pr} may be different from each other. In order to

obtain dense and smoothing surface, the accurate location of the point corresponding with the left point can be achieved by bilinear interpolation according to the distance from that point to the four neighboring block centers in the image I_{pl} . Similarly, when we perform the tracking between the temporal images, we can first get four values by shifting the point with the four shift parameters of the neighboring blocks, and then the four values can be interpolated to obtain the accurate tracking location.

3 Results and Evaluation

In the previous sections, we show how our method can be applied to the reconstructing and tracking of dynamic surface. In the experiment, first, simulated data is used to evaluate the performance of the algorithm. Then,

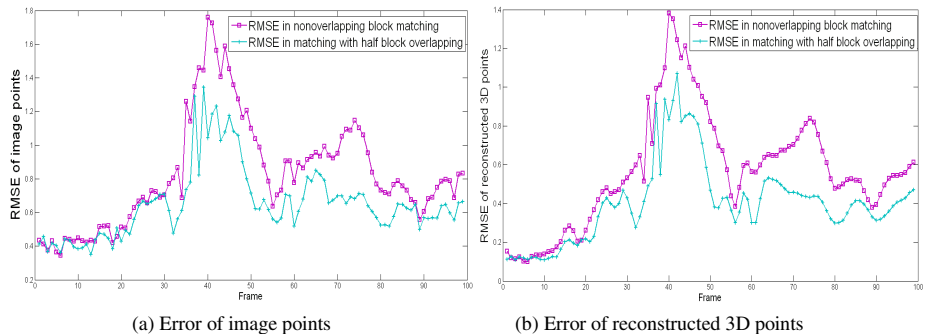


Fig. 3. Error on the ground truth,(a) is the RMSE curve of image points on the ground truth,(b) is the RMSE curve of the reconstructed 3D points on the ground truth. The top curve of the two figures is the RMSE for the matching without block overlapping, and the bottom curve is that of with 50% block overlapping matching.

We validate our approach using real video data acquired by two calibrated cameras with 25fps to demonstrate that it can produce good results for very different kinds of materials, and at the same time, motion field is easily generated by the tracking information.

3.1 Simulated Experiment

To evaluate the performance of our algorithm, the simulated data generated by 3D max is used as ground truth for our first experiment. We produced a stereo image sequence of 100 frames with 800×600 image size in which a piece of cloth is fluttering with different shapes in the breeze, and there is a small movement in previous 20 frames and a sharp movement from 35th frame to 45th frame in the stereo image sequence. At the same time, 400 pairs of vertices of the left and right image sequence and the corresponding spatial 3D points are generated as the benchmark to evaluate the performance of our algorithm.

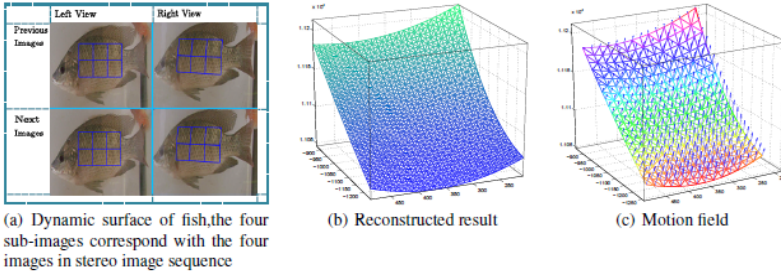


Fig. 4. Reconstruction and tracking of dynamic fish surface

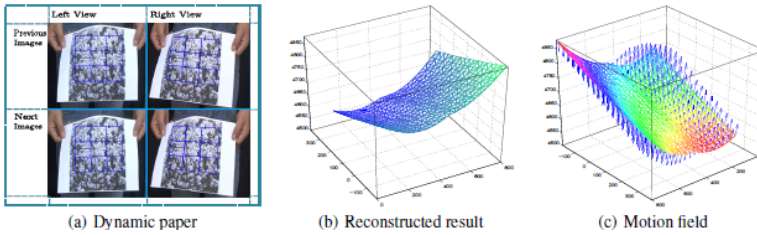


Fig. 5. Reconstruction and tracking of dynamic paper

The block size we used is small: 20×20 in this experiment due to the plenty of texture on the surface. From Fig.3, we can see that when the deformation is smoothed, the reconstructing and tracking result is very good, but for the sharp deformation case, the result will be not perfect.

To solve the problem of sharp deformation, we can overlap the neighboring blocks. Since the block center is the most accurate value in the whole block, and due to the overlapping, the distance between the neighboring block centers will be smaller, which can make the interpolated value in line with the four neighboring block centers in local area closer to the actual value. Theoretically, the reconstructed surface can approximate the ground truth by block overlapping of large size without considering the computational required cost. Fig.3 shows the RMSE of the image points (a) and reconstructed 3D points (b) on ground truth, and the top curve in both (a) and (b) is the RMSE without overlapping, and the bottom one is the RMSE curve with 50% block overlapping, from the figure, we can see that the RMSE is smaller when block overlapping is considered.

3.2 Real Data Experiment

Real dynamic surfaces, such as paper, cloth and fish surface, are used to show the result of reconstruction and tracking by our algorithm.

In our experiment, the size of the images is 1440×1080 , and the block size selected ranges from 30×30 to 50×50 according to the actual size of the surface and the texture. For those surfaces with rich texture, small block size can be used because

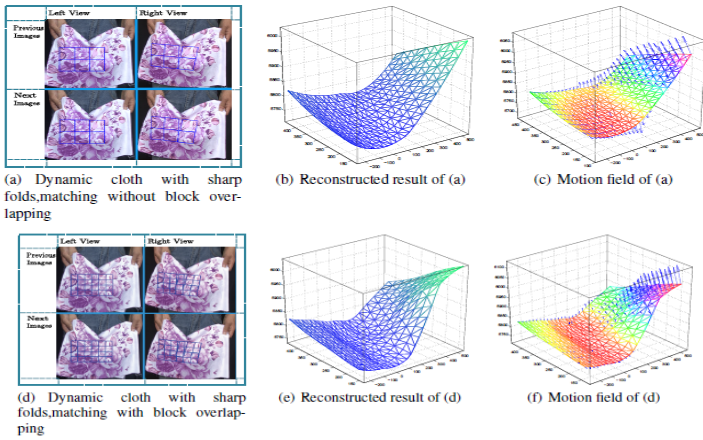


Fig. 6. Reconstruction and tracking of dynamic cloth

the correct solution can be easily achieved, but for those with sparse texture, large block size is needed in order to gain the correct iterative solution.

Due to their different physical properties, the behavior of the surfaces ranges from smooth deformations for the fish surface to sharp folds and creases for the cloth.

However, no parameter tuning was needed to obtain these results with our algorithm. Fig.4 shows the reconstructing and the tracking of dynamic fish surface, (a) is the stereo images in which the top two sub-images are the images of previous frame in left and right image sequence, and the bottom two are those of the next frame, (b) is the reconstructed 3D surface, and (c) is the motion field calculated by the tracking information between the previous and next 3D surface which includes the accurate orientation and distance of the motion of every point in the previous 3D surface. Therefore, this method provides a useful approach to those applications such as analysis of motion rules of the dynamic surface. Fig.5 and Fig.6 a are the results of paper and cloth respectively.

The deformation of dynamic fish surface and paper is usually smoothing as shown in Fig.4 and Fig.5, which deforms smoothly when they are moving, so it is more precise to capture the dynamic data for the motion process. The dynamic process of cloth motion is usually very sharp; in this case, there may exist bigger error in reconstruction and tracking. In order to solve this problem, the neighboring blocks can be overlapped with each other, as mentioned in the previous section. Fig.6 demonstrates the reconstruction of the folded cloth, in which (a) shows the matching without block overlapping, (b) and (c) are the corresponding reconstructed result and the motion field, and (d) illustrates the matching with 50% block overlapping, (e) and (f) are the result and motion field. By comparing (b) and (e), we realized that the overlapping case produces better results.

4 Conclusions

In this essay, we have presented a framework to reconstruct and track the dynamic 3D surface from stereo video by optimizing an energy function. Our approach relies on finding the correspondence among four blocks in a stereo image sequence by obtaining the parameters of transformation, and after all block correspondences are found, all blocks can be integrated into a dense 3D surface by bilinear interpolation. The simulated and real data experiments confirm the performance of our approach. The dense motion field of the 3D surface (which provides an approach to analyze the motion rules of dynamic surface) can be achieved. Since the different block correspondences in one surface can be found independently, it is possible to parallelize block matching in a shared memory environment to speed up the process of reconstruction and tracking. In our future work, we intend to find out the solution to the self-occlusion problem in our framework which is not considered in this essay.

References

1. Ahmed, N., Theobalt, C., Rössl, C., Thrun, S., Seidel, H.: Dense Correspondence Finding for Parametrization-free Animation Reconstruction from Video. In: *Proceedings of Computer Vision and Pattern Recognition* (2008)
2. Baker, S., Matthews, I.: Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision* 56(3), 221–255 (2004)
3. Ballan, L., Cortelazzo, G.: Marker-less motion capture of skinned models in a four camera set-up using optical flow and silhouettes. In: *Int. Symp. on 3DPVT* (2008)
4. Chai, M., Wang, L., Weng, Y., Jin, X., Zhou, K.: Dynamic hair manipulation in images and videos. To appear in *ACM TOG* 32, 4 (2013)
5. De Aguiar, E., Theobalt, C., Stoll, C., Seidel, H.: Marker-less deformable mesh tracking for human shape and motion capture. In: *Proc. CVPR* (2007)
6. Doshi, A., Hilton, A., Starck, J.: An empirical study of non-rigid surface feature matching. In: *Visual Media Production (CVMP 2008)*, 5th European Conference on. pp. 1–10 (2008)
7. Furukawa, Y., Ponce, J.: Dense 3d motion capture for human faces. In: *Proc. CVPR* (2009)
8. Groeger, M., Ortmaier, T., Sepp, W., Hirzinger, G.: Tracking local motion on the beating heart. In: *Proceedings of SPIE*. vol. 4681, p. 233 (2002)
9. Hilsmann, A., Eisert, P.: Tracking deformable surfaces with optical flow in the presence of self occlusions in monocular image sequences. In: *CVPR Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, Anchorage, USA (2008)
10. Huguet, F., Devernay, F.: A variational method for scene flow estimation from stereo sequences. *Research Report 6267*
11. Noce, A., Triboulet, J., Poignet, P., CNRS, M.: Efficient tracking of the heart using texture. In: *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. pp. 4480–4483 (2007)
12. Pekelny, Y., Gotsman, C.: Articulated object reconstruction and markerless motion capture from depth video. In: *Computer Graphics Forum*. vol. 27, pp. 399–408. Citeseer (2008)
13. Richa, R., Poignet, P., Liu, C.: Deformable motion tracking of the heart surface. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008. IROS 2008*. pp. 3997–4003 (2008)

14. Riviere, C., Gangloff, J., De Mathelin, M.: Robotic compensation of biological motion to enhance surgical accuracy. *Proceedings of the IEEE* 94(9), 1705–1716 (2006)
15. Salzmann, M., Hartley, R., Fua, P.: Convex optimization for deformable surface 3-d tracking. In: *Proceedings of IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil (2007)
16. Shen, S., Zheng, Y., Liu, Y.: Deformable Surface Stereo Tracking-by-Detection Using Second Order Cone Programming. In: *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. pp. 1–4 (2008)
17. Shi, J., Chen, Y.Q.: Robust framework for three-dimensional measurement of dynamic deformable surface. *Optical Engineering* 51(6), 063604–1 (2012)
18. Shi, J., Liu, Y., Chen, Y.Q.: Method for three-dimensional measurement of dynamic deformable surfaces. *Journal of Electronic Imaging* 21(3), 033023–1 (2012)
19. Sminchisescu, C.: 3D Human Motion Analysis in Monocular Video: Techniques and Challenges. *COMPUTATIONAL IMAGING AND VISION* 36, 185 (2008)
20. Stoyanov, D., Mylonas, G., Deligianni, F., Darzi, A., Yang, G.: Soft-tissue motion tracking and structure estimation for robotic assisted mis procedures. *LECTURE NOTES IN COMPUTER SCIENCE* 3750, 139 (2005)
21. White, R., Crane, K., Forsyth, D.: Capturing and animating occluded cloth. *ACM Transactions on Graphics (TOG)* 26(3), 34 (2007)
22. Zhou, K., Gong, M., Huang, X., Guo, B.: Highly parallel surface reconstruction. Tech. rep., Tech. Rep. MSR-TR-2008-53, Microsoft Technical Report (2008)

Scene Simulation for a New Type of Armored Equipment Simulator

Guanghai Li, Qiang Liang, Wei Shao, and Xu-dong Fan

Academy of Armored Force Engineering, Beijing 100072

Abstract. The scene simulation is an important content for simulator development. The fidelity of simulation directly affects the effect of the training. In this paper, first we analyse the difficulty of the scene simulation, then combined with the research of a new armored equipment simulator, we analyse and design the scene generation, signal processing and scene display. By doing this, we try to improve the close shot detail, the speed sense of movement and the sense of depth. Testified by the experiments, the training effect has been significantly improved.

Keywords: Armored equipment, simulator, scene simulation and virtual system.

1 Introduction

The simulation is safe, economical and repeatable, by using simulation, we can implement some experiment which can not be carried out in reality. Simulation technology is widely used in the military research and training. In the early 80's of last century, our country began to research the armored equipment simulation training system by using computer simulation technology, and gradually extended to research various types of equipment simulation training system[1]-[3]. Scene simulation is an important part of simulator research the fidelity of scene simulation directly affects the effect of simulation training.

Armored equipment is a kind of ground equipment, the scene simulation of the armored equipment simulator focuses on the simulation of ground scene. Owing to the complexity of the ground characteristics, and the processing difficulty of the movement relative position, in the scene simulation of armored equipment simulator, there are lack of close shot detail, visual feeling weak of equipment movement speed and the lack of object distance, especially in the driving training simulator. The problem directly affect the designated parking training, obstacles training and restrictive roads training in the driving training simulator. In this paper, combined with the development of a armored equipment simulator, we focus on studying the effect of driver scene simulation, try to improve scene detail simulation, the speed scene and depth sense, and we focus on studying scene generation, signal processing and scene display technology.

2 Design of Scene Generation System

2.1 Scene Software

The scene is created by using advanced engine software, the operating system is Win7. In the scene, the vegetation, roads and building is presented by 3D model.

(1) Virtual battlefield environment: man-made features include restrictive roads, obstacles, targets, barbed wire, and anti-tank pyramids, natural features include houses, roads, vegetation, bridges and rivers, natural climate include cloudy, sunny, rain, snow and fog.

(2) The 3D model of weapon: we establish 3D model of armored equipment.

(3) The battlefield effect library includes driving imprinting, dust, gun muzzle flash, surface of water, ground explosion, dynamic crater effect, equipment damage and burning effect, infrared thermal image effect, the dynamic damage effect of the houses, bridges and trees in natural scene.

(4) The battlefield sound effect library includes the sound of rain, the sound of wind, the sound of explosion, the sound of fire outside the vehicle, the sound of vehicle running and the internal sound of equipment and weapon.

2.2 Hardware Platform

According to the theory of visual perception, the visual angle of normal (1.5) eye is 2 points, the human eye is unable to identify the object which visual angle is less than 2 points.

Therefore, if the single channel scene resolution is 1920*1080, total resolution of the three channel fusion is 5134*1080, when the maximum horizontal viewing angle is 160 degrees, visual angle of per pixel is $160/5134*60=1.87$ points, this result can meet the need of human eye observation, under this resolution the human eye can not feel the existence of a single pixel.

The computer hardware should meet the high resolution graphical calculation, also should ensure the synchronous output of three channel scene. Therefore there is a higher requirement for the computer hardware structure. In order to reduce the number of computer, also reduce network traffic and improve the system reliability, we use three sets of high performance HP graphics workstation as a graphic computer, every computer is equipped with a AMD W7000 video card, and every computer outputs three channels of high definition video signal.

2.3 Analysis and Design of Surface Texture

As shown in Figure 1, when the driver is observing the front ground through the periscope, according to the nearest distance observed using periscope (observation lower bound is 10 meters) and the view angle of virtual camera(100 degrees) (depress the driver's seat) , we can calculate the width of observed ground.

$$W = 2 * d * \tan(100/2) . \quad (1)$$

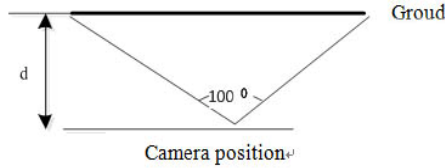


Fig. 1. The analysis diagram of view resolution

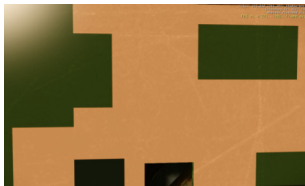
When d is 10 meters and w is 24 meters, the width of the ground is about entire 62.5% of the projection screen, there is about physical pixel 3200 in the horizontal, the texture pixel density is:

$$\text{Density} = 3200/w = 133 . \tag{2}$$

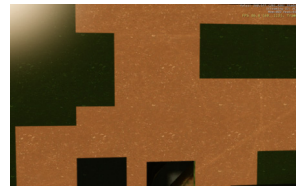
If per meter has 133 pixels, it can meet the need of driver's eyes observation. In order to meet the need of close shot visual effect, we can take the following measures.

(1) Using high resolution ground texture: surface texture resolution of 10 meters road is 2048 x 2048, 1 meter has 204 pixels.

(2) We use the method of diffuse texture overlay detail texture for the tank body first armor, this method can improve the surface details and texture when observing closely. The effect is shown in figure 2.



(1) without using overlay texture



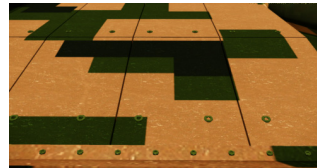
(2) using the overlay detail texture

Fig. 2. The effect contrast of detail texture

(3) Instead of texture description, we construct the model for body first armor plate again, including the model of reactive armor and nut. Effect contrast is shown in figure 3.



(1) The texture effect



(2) The entity modeling effect

Fig. 3. The effect contrast of texture and entity modeling

2.4 The Sense Simulation of Movement Speed and Depth

The sense of speed is important to improve the fidelity of driving simulation training. Comparing with the actual driving, the sense of speed in virtual scene also needs reference[4]. We can see from Figure 4, there are a lot of trees on both sides of the road, the ground grid and not flat state can enhance the speed sense of tank driver.



Fig. 4. The road simulation effect having reference

The sense of distance influence the driver to determine the distance accurately, especially designated parking, obstacles and restrictions training is obvious. The display produces the plane, so the sense of distance is usually realized through the stereo imaging and virtual way. By evaluating the advantages and disadvantages of various methods, the system software adopt hierarchical modeling mode, it can show the effect of the different prominence, we use the positive projection image system, then we can generate wide field angle and high resolution scene.

3 The Design of Visual Display System

3.1 The Scheme of Virtual System

Visual display system is an important part of the virtual battlefield environment, its main function is to provide the graphics to the observer using display device or equipment. The depth is an important index of visual fidelity degree, virtual display and stereo implementation can provide realistic depth. WIDE (wide-angle infinity display equipment) system is a commonly used off-axis virtual image display system, it is composed by projector, spherical mirror and scattering screen [5-6]. According to the imaging way on the scattering screen, it can be divided into positive projection and back projection. The principle diagram of positive projection image display system is shown in figure 5.

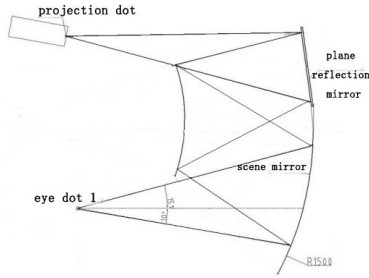


Fig. 5. The principle diagram of front projection image display optical system

The projector output graphics is imaged on the scattering screen through the plane imaging mirror, the scattering screen is located in the 1 times plane of spherical mirror, the graphics is imaged to infinity reflected by the spherical mirror. Because reflected only 3 times, the loss of brightness is small, the imaging effect is good.

3.2 The Construction of Horizontal Wide Field Angle

Restricted by the production technics, producing the wide area virtual mirror is difficult. The system adopts the mode of "3+2", namely the combination of 3 blocks of large virtual mirror and 2 blocks of small virtual mirror, the center angle of large virtual mirror is 35 degrees, the center angle of small virtual mirror is half of the large mirror, the overall angle is 140 degrees, the overall is symmetrical. The eye point is on the front position, viewing angle is 160 degrees. The cost is relatively high, splicing method is shown in figure 6.

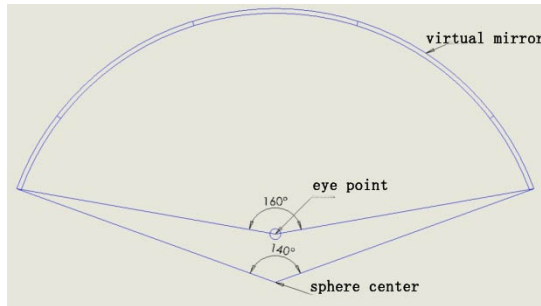


Fig. 6. The splicing scheme of virtual mirror

4 The Transmission Fusion of Video Signal

In the simulator, there has a lot of passenger scene windows, and the scene display terminal is distant from scene computer, if we use the traditional video cable transmission, it will cause the interference between the signals, also affect the quality of the image, and the motion of the platform will cause damage of the cable or connector, the reliability of system will be greatly reduced.

Video signal transmission scheme is shown in figure 7. The output video signal of graph workstation is converted by video optical transmitter, it is restored by optical receiver through the optical fiber, the graphic signal is integrated and corrected by fusion system, at last it is exported to the three projection equipment. This scheme has strong anti-interference ability, it is suitable for electromagnetic signal complex environment, if we use the good fusion system, we can ensure the signal transmission delay time is less than 40ms, it can meet the need of simulation training.

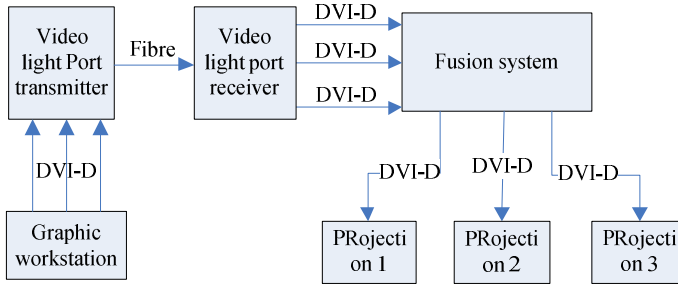


Fig. 7. The scheme of video transmission connection

5 Application Effect

The Scene simulation is one of the important contents of simulator. The scene simulation is also difficult, especially when (because?) the ground state is complex, terrain object and vegetation are complicated. How to realize the high fidelity of the simulation effect has been the goal of struggle. Combined with the new simulator development task, we focus on enhancing the close detail, the sense of movement speed, the sense of depth in the scene simulation. The experiment proves that the method is feasible, and the effect is obvious, especially the method fits the armored equipment driving simulation training.

References

1. Pu, H., Li, W., Zhao, H.: Integrated Model Construction and Simplification Methods for Web 3D road. *Journal of Computer Applications* 33(02) (2013)
2. GJB4512-2002. The general specification of armored vehicle driving simulator
3. Yu, J., Cong, M., Wang, Z., Cao, B.: The motion simulation arithmetic and implementation of tank driving simulator. *The Computer Simulation* 24(7), 305–308 (2007)
4. Ji, L.-E., Sun, X.-Y., Guo, W.-S.: Research on Real-time Virtual Simulation of Underwater Dynamics Environment Based on Vortex. *Journal of System Simulation* 9 (2013)
5. Huang, A.-X.: The design of virtual battlefield. National Defence Industry Press (2007)
6. Xu, J., Wang, S.-P., Li, T.: Research on Realization of Surface Ship Large Simulation Training System's Key Technologie. *Journal of System Simulation* 8 (2013)

Study on the Key Technique of the Hill Shading Virtual Roaming System Based on XNA

Hesong Lu, Zaijiang Tang, Qiang Liang, and Wei Shao

Simulation Center, Academy of Armored Force Engineering, No. 21 Dujiakan,
Fengtai District, Beijing, China
ww2_db@163.com

Abstract. It is helpful for people to understand the undulating topography with hill shading map. In this essay, we focus on how to shade the terrain blocks with XNA Game Studio3.1 by using the global DEM data from geographic information public service platform set as the data source. Including LOD structure, constructing a TriangleStrip terrain block model, applying HSV color model to color vertices, using XNA default lighting model and adjust relevant parameters to generate shading effect. Finally, a global height-color mapping table designed was offered, the represented effects of prototype system were also showed.

Keywords: XNA, DEM, hill shading, LOD, HSV color model and roaming system.

1 Introduction

The event of Malaysia Airlines MH370 aircraft lost made us realize that we have limited understanding of world geography. Digital Earth and other related public GIS products such as Google Earth in virtual reality did very well, but the lack of hill shading terrain roaming function still needs to be developed. The hill shading method is crucial to produce the three-dimensional effect of topography on the map [1]. Making hill shading map is one of the basic functions in GIS, but they lack the large terrain roaming capabilities or must be online to use. TIANDITU website v2.0 supports browse hill shading map, but does not support the use of 3-D hill shading terrain roaming.

The 3D terrain visualization technology is widely used not only in GIS and virtual environment simulation but also in 3D computer games. The XNA Game Studio is a game development kit designed by Microsoft for independent game developers. Compared to DirectX and OpenGL, XNA meets the functional requirements of the system and is more efficient. The geographic coordinate system of the DEM data used in system is WGS-84. It is a right-handed coordinate system, and the default coordinate system in XNA is right-handed. The prototype system developed the XNA Game Studio 3.1.

The rest of this essay is organized in the following ways: Section 2 introduces the LOD structure in brief; Section 3 contains the description of the terrain block model;

in Section 4, the method of shading terrain block model in XNA is explicated; in Section 5, the shading effects of the roaming system prototype are offered and the conclusions are discussed in the last Section.

2 LOD Structure

It cannot do without DEM data to create hill shading effect [2]. The roaming system uses DEM data to download geographic information from the platform of common services, such as CGIAR-CSI, NASA, etc. These data sets include the DEM data of six kinds of resolution. They are SRTM 90m V4.1, ASTER GDEM and re-sampled SRTM data to 250m, 500m and 1km.

These DEM data take up a lot of storage space. The size of SRTM 500m DEM ASCII file format data after decompression is more than 10G. These data can't be pre-loaded during the system initialization in the main memory. The LOD technique, which doesn't represent the main focus of this essay, is commonly used to build a large-scale terrain roaming system in engineering. Only the method described here is used in the prototype. The prototype designed to use the static LOD. Select the LOD level based on the distance between the viewpoint and terrain surfaces. The DEM data of each level was split into a group of small terrain block files. Each block file has 301×301 height data. The choice of this size was decided to keep the accuracy of the original DEM data, yet without the requirement for data interpolation. The data exchange format of terrain block file is ArcInfo ASCII(described in reference[3]). The terrain block files at the same level were named by the unified coding.

It was designed to add three levels LOD data to further simplify the DEM data. These data were re-sampled based on SRTM 1km ASCII data, that signed in the following table(Table 1) by marked * in LOD level column. The resolution of other levels remains.

Table 1. Resolution of DEM data at each LOD level

LOD level	Data type	resolution of DEM (°)
1*	SRTM 12km	0.099999999996
2*	SRTM 6km	0.049999999998
3*	SRTM 2km	0.016666666666
4	SRTM 1km	0.008333333333
5	SRTM 500m	0.004166666666
6	SRTM 250m	0.002083333330
7	SRTM 90m DEM	0.000833333333
8	ASTER GDEM	0.000277777778

3 Terrain Block Model

3.1 Defining the Vertices

The prototype system uses color shading. While creating hill shading effect requires the use of lighting model in XNA. So the vertex structure of terrain block should also

include vertex normal and vertex color in addition to vertex position. The class of vertex should be defined by user in XNA.

The structure of the DEM data used is regular grid. It is suitable to load in array. Filling the vertex array occurs in system initialization and updating of the roaming camera's view point. After vertex position data read from the small terrain block file by `FileStream` instance, it needs to be transformed from the WGS-84 geocentric geographic coordinates (B, L, H) to geocentric rectangular space coordinates (X, Y, Z). The conversion formula can be found in reference [4].

3.2 Defining the Indices

Using a `TriangleStrip` primitive type can save more memory and bandwidth than using a `TriangleList`. A method that generates indices defining triangles as a `TriangleStrip` for a terrain based on a grid is described in reference [5]. Figure 1 shows the order of defining the triangle indices. In reference to the basis of his thinking, the order of defining the triangle indices designed in prototype system was shown as in Figure 2. After added index pointing to vertex $(3*w-1)$, add next index pointing to the same vertex again. In fact there are four invisible triangles between the normal triangle based on vertex $(2*w-2)$, vertex $(w-1)$ and vertex $(2*w-1)$ and the normal triangle based on vertex $(3*w-1)$, vertex $(2*w-1)$ and vertex $(3*w-2)$. Thus the culling order is not changed.

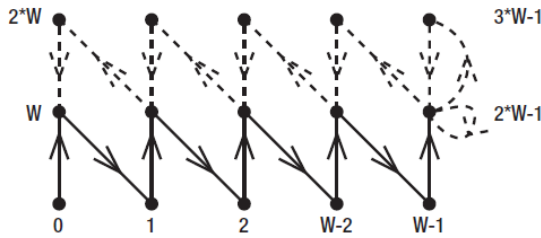


Fig. 1. "Rendering your terrain correctly as a TriangleStrip" in reference [5]

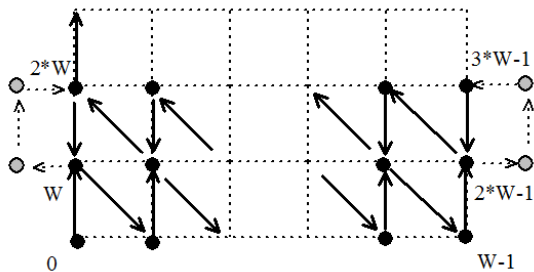


Fig. 2. The order of defining the triangle indices designed in prototype system

Because the size of each terrain block file is the same 301×301, the array of vertices indices need create in system initialization phase once.

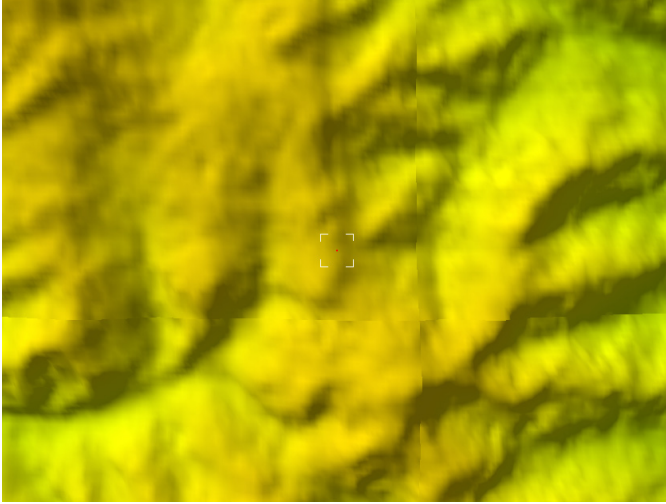


Fig. 3. The lighting effect is not smooth in the seam

3.3 Terrain Splicing

The normals of vertices can be calculated based on the position of the vertices and the indices defining triangles. A method of calculating vertex normal can be found in reference [5]. But when a group of terrain blocks are splicing, because the vertices at each block boundary miss parts of information about normal, the lighting effect is

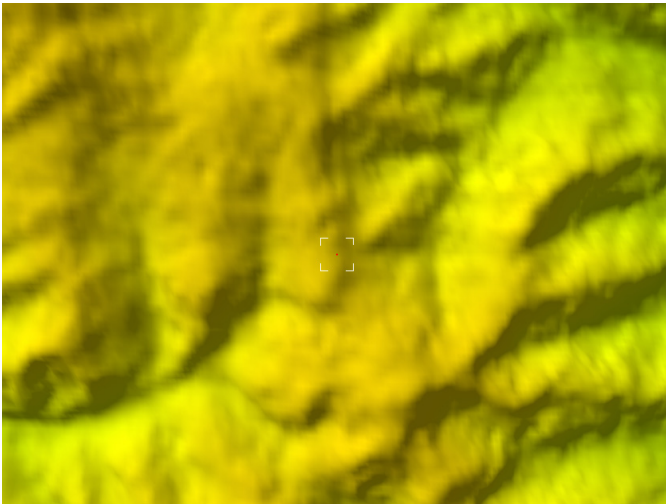


Fig. 4. The lighting effect after terrain splicing

not smooth in the seam. The effect was shown as in Figure 3. It was designed to solve this problem that merging the normal of the same vertex of on the adjacent side in adjacent block, before the boundary vertices normals were normalized. Then normalize normals. The effect after terrain splicing was as shown in Figure 4.

4 Shading Terrain Blocks

The application of the hill shading uses color shading. The value of the vertex color needs a frame of reference. Common method is to establish height-color mapping table. To establish the mapping table with reference to the traditional color habits on one hand: blue for ocean, green-brown for land, white for high mountains; on the other hand in accordance with “the higher the brighter”. The height interval was designed between -11100m and 8900m. Because the interval is so long, the red color and grey color were added.

The value of these DEM data on the parts of ocean was filled with an invalid value -9999 or 0. Although some parts of the data is invalid, but consider the possible future expansion of the system, still the zone of blue reserved to represent ocean. After setting up the mapping color on key height, the color mapping of different height can be calculated using linear interpolation by using the table as reference.

As the RGB color model of the computer is not intuitive, the HSV color model was selected to represent colors. The conversion algorithm of HSV value to RGB value can be found in reference [6]. As shown in the Figure 5, the direction of the arrow represents the height of the system from low to high, the corresponding color tint gradient sequence.

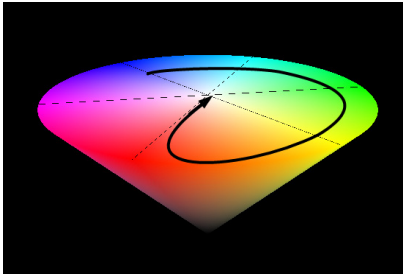


Fig. 5. HSV color model

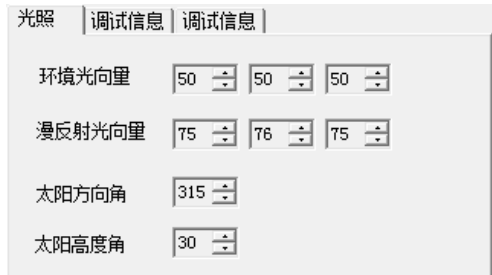


Fig. 6. Interface of Lighting parameters setting

XNA supports flexible programmable pipeline. The system uses an instance of the `BasicEffect` class in XNA to achieve basic lighting effects on terrain model.

There are four main factors to control lighting effect in XNA. They are azimuths of the sun, altitude of the sun, color of ambient light, color of diffuse light. As the hill shading effects is produced due to illusion and psychological association [1]. Because of the existence of the pseudoscopic effect [7], the four parameters are opened to the user. Figure 6 shows the interface of setting these parameters.

This example code shows the key parameters setting of the `BasicEffect` instance in XNA:

```
basicEffect.VertexColorEnabled = true;
basicEffect.LightingEnabled = true;
basicEffect.PreferPerPixelLighting = true;
basicEffect.AmbientLightColor=mainform.AmbientLightColor;
basicEffect.DirectionallLight0.Enabled = true;
basicEffect.DiffuseColor = mainform.DiffuseLightColor;
Vector3 terrainPos =
GetTerrainCenterPos(cameraSat.PlatLong,cameraSat.PlatLat);
Vector3 lightDirection = GetSunDir(terrainPos);
basicEffect.DirectionallLight0.Direction = lightDirection;
basicEffect.DirectionallLight0.DiffuseColor =
Color.White.ToVector3();
```

Figure 7-8 shows the contrast effects between lighting off and on.

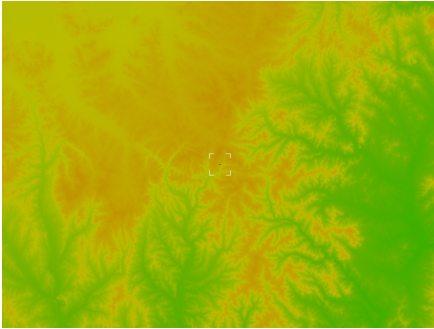


Fig. 7. Lighting off effect (LOD at level 7)

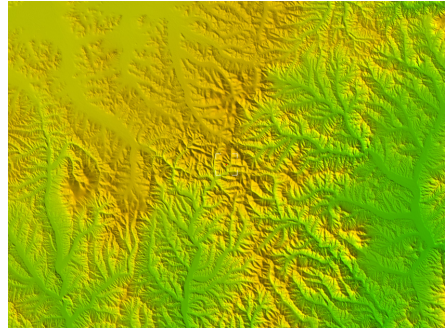


Fig. 8. Lighting on effect(LOD at level 7)
azimuths of the sun =225° altitude of the sun =30°

5 Representing Effects of Prototype System

5.1 Running Environment

CPU: Intel(R) Core(TM) i7-2670QM CPU @2.20GHz 8core;
Main Memory: 8G;
Graphics Card: NVIDIA GeForce GTX 670M;
OS: Windows7 64bit.

5.2 Some Hill Shading Effects

The experimental data of height-color mapping table as shown in the following table (Table 2).

Table 2. Height-color mapping table

Height(m)	color value	height(m)	color value	height(m)	color value
8900	0,0,1	1600	52,0.99,1	10	130,1,0.3
6400	359,0.02,1	1000	84,1,0.9	5	136,1,0.25
5000	295,0.13,0.62	600	100,1,0.9	-5	180,0.5,0.8
4000	0,1,0.83	100	110,1,0.6	-30	200,1,1
3500	16,0.95,0.82	50	120,1,0.5	-100	240,1,1
3000	24,0.75,0.88	30	126,1,0.4	-11100	250,0.5,0.5
2400	43,1,1				

Figure 9-13 is a group of representing effects of hill shading by prototype roaming system with the parameters in above height-color mapping table.

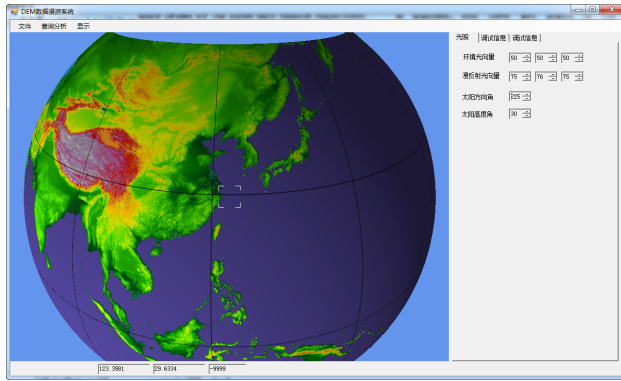


Fig. 9. Overview of prototype roaming system(LOD at Level 1)

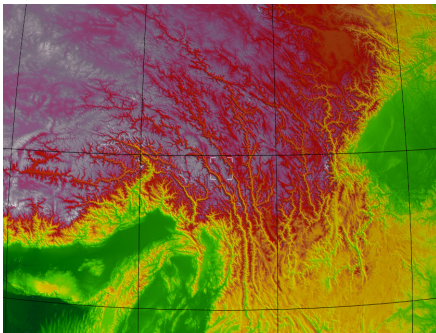


Fig. 10. Shading effect (LOD at level 3
azimuths of the sun =225°
altitude of the sun =30°

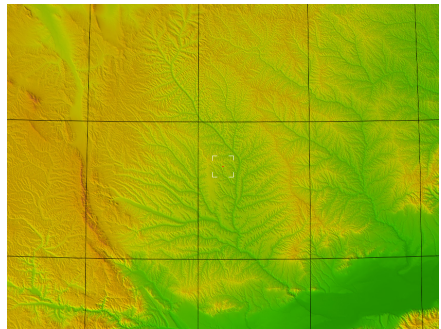


Fig. 11. Shading effect(LOD at level 5)
azimuths of the sun =225°
altitude of the sun =30°

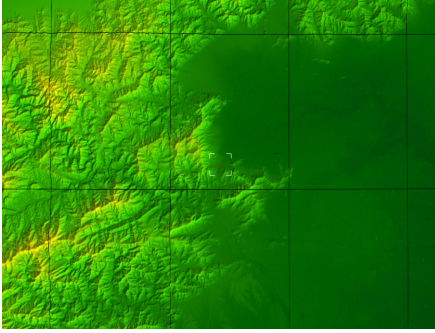


Fig. 12. Shading effect (LOD at level 7)
azimuths of the sun = 135°
altitude of the sun = 30°

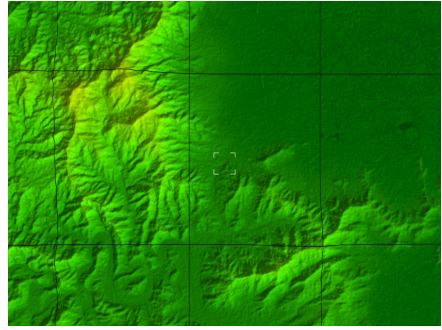


Fig. 13. Shading effect(LOD at level 8)
azimuths of the sun = 135°
altitude of the sun = 30°

6 Conclusions

In today's hardware conditions, the use of static LOD can meet real-time. While drawing 21 terrain blocks, the frame rate of prototype system can maintain at 29fps. The prototype experiment shows that the application of the visualization technique feasibility is based on XNA.

At present, the shortcoming in the prototype system is that the scheduling algorithm for terrain blocks is not optimized. If the frame rate was improved, system can add other functions, such as terrain analysis.

In our further research, we will use the self-defined lighting model by HLSL, and we will study the relationship of lighting parameters.

Acknowledgments. Thanks to NASA and CGIAR-CSI! Without these public DEM data provided by them, these ideas cannot be achieved! Thanks to the physics group of Shanghai Eighth Middle School! Their annotation and experience made the learning of XNA easy!

References

1. Wang, J., Sun, Q., Wang, G.: The principles and methods of cartography. 地图学原理与方法. Science Press, Beijing(2011)
2. Li, Z., Zhu, Q.: Digital elevation model. 数字高程模型. Wuhan University Press, Wuhan (2001)
3. Zhou, Q., Liu, X.: Digital terrain analysis. 数字地形分析. Science Press, Beijing (2006)
4. Li, Z., Wei, E., Wang, Z.: Space Geodesy. 空间大地测量学. Wuhan University Press, Wuhan (2010)
5. Grootjans, R.: XNA 3.0 Game Programming Recipes: A Problem-Solution Approach. Apress, Berkeley (2009)

6. Zuo, F., Wan, J., Liu, H.: Visual C++ digital image processing development and programming practice. Visual C++数字图像处理开发入门与编程实践. Publishing House of Electronics Industry, Beijing (2008)
7. Chang, K.-T.: Introduction to Geographic Information Systems. In: Chen, J., Zhang, X.(trans.) 地理信息系统导论. Science Press, Beijing(2010)
8. Hu, J., Li, L., Li, X.: Reseach on computer hillshading based on directX. Acta Geodaetica Cartographica Sinica 4, 20–22 (2004)

Study on MMO in Visual Simulation Application

Liang Qiang, Fan Rui, Xu Renjie, and Du Jun

Academy of Armored Forces Engineering, Beijing
lqcxjpp@wo.com.cn

Abstract. MMO is the network communication mechanism that has been applied in large scale network game. The related technology of MMO is quite mature, and it has relative hierarchy. However, in the DIS visual simulation filed, with the enlargement of simulation scale and the increasing of node, the load of network is much greater than before, and it will decrease the real-time of the real-time rendering at every node. In this essay, we will provide analysis of the characteristic of data exchange in visual simulation, and explain the technical advantage of MMO on communication mechanism. Then we will present our studies on how to effectively reduce the network load, enhance the virtual scene real-time rendering the visual simulation, give a preliminary solution and conduct several proving experiment. Based on above mentioned research work, we built the foundation on real-time data network interaction environment that can be applied to large scale DIS visual simulation and be easily extended.

Keywords: MMO, DIS, Visual Simulation, Real-time and network load.

1 Introduction

Currently whether in the military DIS simulation training system or the application of large scale network game, in order to get the best network simulation training effect and the fluent visual game experience, the real time data interaction between every two network nodes is very important and pivotal. There are many related technology realization mechanisms and methods that DIS and HLA are applied in military distributed interaction simulation, and MMO (Massively Multiplayer Online) is the principal way in the network game. As author's experience from some military distributed interaction simulation training system indicates, when the simulation scale enlarges to 150~200 simulation node or more, it will appear rather serious transmitting delay and data lose, it can't satisfy the real time need of the visual simulation of virtual scene rendering, therefore it would affect the simulation training result. Relatively, MMO network game always has more than 10 thousands or even several 100 thousands players online at the same time, the direct experience of network transmitting real time of related game data can be likely ensured.

In this essay, we analyze the causes that can decrease real time in military distributed interaction visual simulation, study the application of MMO mechanism in DIS

for reducing network load and flow, improve the real-time visual simulation, and explain our idea and method of preliminary solution.

2 DIS Visual Simulation and MMO

With the improvement in science and technology, current combat is becoming informatization and many dimensions; the traditional training and practice mode does not fit the requirement of modern war. The computer simulation has some advantages, such as low risk, low investment, and no influence of time and environment and so on; and it replaced the traditional way. However, one single simulator has solved the problem of training of a single member, it can't fit the needs of high level training task such as cooperation and unit tactic. For example, to train the best commander, best driver and best shooter at the same time, the fighting capacity of this combination may be weak; in addition, one single vehicle has powerful fighting capacity, but when they form a platoon or company, the fighting capacity may not be powerful. The DIS just deal with this problem[1]. Among them, the key to the training experience in direct vision is the virtual battlefield scene simulation, fluent.

In 1993 started the SIMNET research plan which was made by USA DARPA and US Army together. Until now, DIS goes through 4 main phases: development and application of SIMNET, development and application of DIS and ALSP and development and application of HLA after 1995. The USA was the first country to promote DIS/HLA technology study. Under the leading of DMSO, many crews from industry, research departments, colleges and troops have allied together to focus on this technology, and make great progress. The DIS of research has been applied, and accomplished many DIS and ALSP engineering project based on virtual simulation. Related agreement and standard have been completed or are in the standardizing procedure[2].

2.1 DIS and HLA

When discussing DIS, the following concepts will be used: simulation entity, simulation node, simulation application, simulation management computer, simulation practice and simulation host machine [1].

There are many simulation nodes in one DIS network. Each simulation node can be a simulation host machine, or a network switch device. One or more simulation application, which can interact with each other, makes up a simulation practice. The simulation application in one simulation practice would share one practice identifier. Each simulation application is responsible for maintaining the state of one or more simulation entity; and the simulation entity is one unit in simulation environment. If the simulation host machine has the simulation management software, this host machine becomes a simulation management computer that is responsible for completing part or overall simulation management function.

The basic concept of DIS, such simulation application, simulation entity, and simulation host machine, etc., is still used by HLA. However, some concept is described with other name and intension of some has been expanded in HLA. HLA also brings some new concepts [3] as explained in following paragraph:

It still uses conception of simulation application, but it is called federation member in great majority case. The conception of simulation practice is replaced by federation. The federation is composed of many federation members, namely many simulation applications. Run Time Infrastructure (RTI) is a new conception in HLA. From physical view, it is a kind of software that distributes in every simulation host machine of HLA, like a distributed operation system. Every simulation application informs RTI what data it will send or need by communicating with local RTI; then the RTI is responsible for communicating with other simulation application. So the information exchange can be realized.

2.2 MMO

MMO is the network communication technology of large scale network game. MMOG (Massively Multiplayer Online Game) system model consisting of a centralized structure to a distributed architecture development is an inevitable trend[4]. In large scale many player online game, it needs transmitting posture information such as position and speed between players. It has a greater communication flow. In order to solve this problem, MMO designs several mechanisms to decrease the communication flow and reduce the network load. Generally, this mechanism is called MMO communication mechanism. With the use this mechanism, the network load which was produced by transmitting above information between players in game scene can be effectively reduced. According to characteristic of the network game, like big communication flow, high need of real time, it usually uses UDP mode on agreement, but the communication architecture is different between games based on its own requirement and characteristic. Currently, the P2P structure is a widely using structure. Paper also uses UDP agreement and P2P structure in realization.

Peer-to-Peer computing is a study hotspot recently. The characteristic of the peer-to-peer computing is that it deals better with server computing bottleneck problems of C/S mode, the computing capability of system is decided by the sum of all node computing capability. Since every peer represents an equal node, no one has any privilege, therefore, there is no total data control ability. The virtual source of player is stored in local node, but the safety can't be well ensured. In theory, with the joining of more and more peer, the system computing capability will be increased and the availability will be enhanced, but the communication volume of the entire system will be increased by n^2 , of course, we can use some technology like interest management to make improvement [5].

3 Application Analysis of MMO Mechanism in DIS Visual Simulation

3.1 Requirement of Real Time

The Real time is one basic characteristic of DIS visual simulation. It requires transmission of a big size of data such as words, image, sound and position in time in the

system simulation process between every simulation node. In addition, it should make sure of space consistency in whole, that means all simulation nodes are running on the same battlefield terrain and the environment database, Real time rendering of virtual scene of each node according to the received simulation entity space position and attitude data. According to the principle of persistence of vision, to achieve human feeling relatively smooth on visual effect, the virtual scene rendering frame rate should at least reach 24 frame / sec.

3.2 Current Problem

Because of the increasing of simulation node, the main problem in DIS is that the real time communication is not ideal enough. Its main cause is that DIS just supports a single broadcast data communication which is based on non-connect. Its features are:

- (1) It cannot ensure the data sending and receiving accurately.
- (2) Data sending uses the way of non-connect broadcast to send interaction information on network ignoring whether the other simulation node needs these information.
- (3) Other on line nodes can only passively receive interaction information from network, then decides how to deal with these information.

The fault of this way is that the receiver should receive all information from network first, then judge whether the information affects itself, process valuable information and delete useless information. The study shows that information interaction quantity between simulation node will increase exponentially with the increasing of the entity number of simulation.

In summary, network bandwidth is occupied by mass unnecessary data transmitting, this is the main reason that causes the unsatisfying real time in DIS. DIS and HLA are short of a mechanism to confirm the communication target. If we can find a way to let every node to communication with related node, the communication flow will be effectively decreased, and we can aim at the improving of real time in simulation.

3.3 Solution of MMO in DIS Visual Simulation

The network load brought by the great communication flow is the cause of the delayed communication. The key problem is how to reduce the network load. In DIS, so many entities need to communicate with each others, we can consider this problem from two aspects: first, should these entities need to communicate? Second, how can we reduce the communication flow during interaction? Based on those considerations, we should concentrate on the effect of decreasing communication flow, and improve the network communication and the virtual scene real-time rendering.

- (1) Using the visual field partition interaction range to confirm the interaction objects of every entity

The military simulation is different from network game. In network game, the plays are regarded as equal distribution, and have the same visual field. So, we can use the nine-rectangle-grid mode and use the space dividing technology to confirm the interaction objects to reduce the interaction flow. But in DIS, different simulation entity

has different visual field, and its distribution is not equal. Therefore, we can use the method of visual field partition interaction range to divide space, namely every entity node has its own visual field space and has its own interaction objects because the entity doesn't need to communicate with all nodes in the space.

We assume that there are three entity node, tank, armored car and soldier. If we don't consider the interaction range, all possible interaction would be as in figure 1.

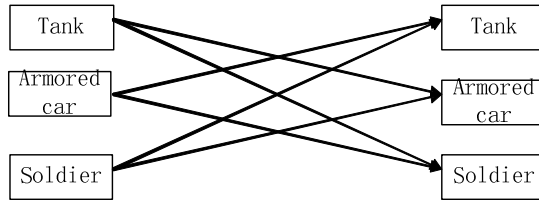


Fig. 1. Possible Interaction of Entities

We can use the visual field partition method to confirm the interaction range as figure 2 shows. The radiuses of visual field ranges are 2500m, 1800m, 1000m, among them, tanks and armored vehicles were observed by sight certain magnification, the soldiers were observed only by the naked eye, and they are moving from faraway place. When soldier moves to the tank orientation and enters into the tank visual field, the distance between them is 2500m, but the tank is 1500m away from the soldier's furthest visual field border at this time, so, we just need to transmit the soldier's entity node simulation information to the tank, and the tank doesn't need to transmit the information to the soldier, as figure 2 shows.

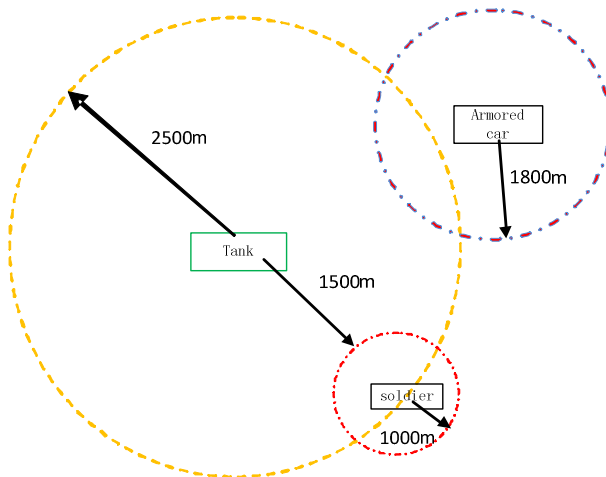


Fig. 2. Interaction range of visual field partition

Virtual scene corresponding to the display as shown in figure 3:

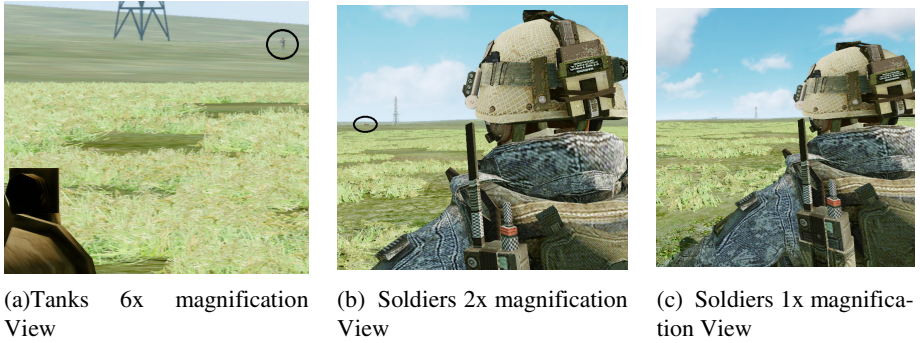


Fig. 3. Tanks and soldiers each 3D view display

Interaction of entities will change accordingly at this time, as figure 4 showing.

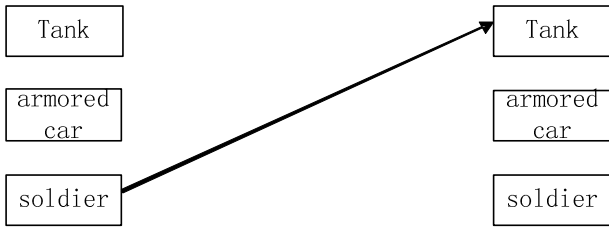


Fig. 4. Entities interaction after Interaction range of visual field partition

If it needs to transmit 20 data packages in one interaction between this three entities (the entity doesn't need to interact with itself), there are 120 packages according to the mode of figure 1 in one data interaction, and just 20 packages according to the mode of figure 4. Therefore, by using mode of visual field partition interaction range, we can obviously reduce the simulation data flow on network. Furthermore, the entity scale is bigger and the effect is more evident.

(2) Using Dead Reckoning technology to decrease the communication frequency

During the network peak time, network often works with delay and jam. If players always wait for the updating message back from the server in large multi-people game, then player would loose the interest for the game. Under this situation, we should use some measures to reduce the negative effect which were brought by the network delay to remove network delay result. Dead Reckoning is a method to ensure the fluency of game under bad network status. Dead Reckoning means that if updating data from the server does not arrive now, client reckons next state based on the current state, of course these state should be continued. The forecasting position of players is just the simple case. In this case, the client can conserve the resent N position and its related time. When $N > 3$, we can use the twice function to make interpolation computing, so that it can forecast the position at a new time by twice equation. If $N = 3$, there are three position in client, the sequence according to time is P_0, P_1, P_2 , the corresponding time is T_0, T_1, T_2 . By

using twice interpolation function to compute position P3 at time T3, after the computing, we can forecast position of T4 according to recent three time T1, T2, T3.

When the network status goes better and the updating data arrives, the old position data should be thrown away, and we should build twice function based on the real data. If there is deviation between forecasting data and data from the server, the game will bring a “jump”, namely jump to a new position from the wrong position. The solution is to smooth this “jump” effectively by gradually using the closing interim algorithm [6]. The study shows that Dead Reckoning technology can improve the game fluency distinctly if network delay is not very serious.

Dead Reckoning is also called DR algorithm. The mathematical formula of DR algorithm is the out push formula that describes the going forward with time. The principle equals to forward comprehensive method derived with time. Supposing that there are entities A, B and C in the three-dimensional Cartesian coordinate system. When state of A has changed, we compare its real state with value from DR model: if interpolation of the two is bigger than the earlier setting threshold value, A will send its updating entity state, and at this time, B gets state information of A by interpolation which is from using DR model by making use of the nearest state information. When entity receives new external updating state information, it will conduct new interpolation computing according to current state information.

The Effect of DR algorithm can be measured mainly by communication times between entities and reckoning deviation. The major factors to influence the result of DR algorithm include: order of DR algorithm, threshold value and transmitting delay between entities.

In the system which uses DR algorithm, when one entity receives updating state of another entity, if it uses this data to update position, the image will appear phenomenon of jumping or be not continued. It will affect the simulation result, so it needs to be smoothed. In order to get a better smoothed algorithm and satisfying requirement of visual scene display, we should consider the following three factors:

- a. The continuity of visual scene: visual scene do not appear “jump” phenomenon, namely the position, speed and direction of entity don’t change suddenly.
- b. The veracity of position: deviation between object position in visual scene and its real position should be as small as possible.
- c. The veracity of action: speed and acceleration of entity in visual scene should reflect action of real object.

The smoothed process technology in Dead Reckoning mainly is polynomial function to complete interpolation computing between two points, so we can get smoothed result. Different interpolation method has different effect, the easiest way is the linear interpolation.

Receiver selects Tsm as smoothed time, and uses new recursion model to calculate position Pf at smoothed end point tf. The smoothed beginning point is the last point Ps, which is calculated by old recursion model before receiving updating data. The line PsPf is smoothed track that transits from old DR model to new DR model. Because the track is line, so we call it linear smooth algorithm. In figure 5, $P_f = P_{new} + V_{new} * T_{sm}$, Pf is the smoothed end point, Pnew is the updating position through delay compensation, Vnew is the speed value at updating, Tsm is the smoothed time [7].

In a word, supposing that at time t , entity node 2 discovers object entity node 1 (namely 1 goes into visual field of 2), the entity 1 doesn't discover the entity 2. From this time, entity 1 should give own state information to entity 2. At the entity 2 position, the visual retention of human is 24 frames per second, so it should transport 24 times position information that the image display which we are observing don't be get stuck. If entity 1 runs from point m to point n using 4 seconds, it needs to transport 96 times position information. By using DR algorithm, under the effect of forecast mechanism, if it needs to transport 10 times state information per second, so the entity 1 just needs to transport 40 times state information during its moving procedure. The communication time is $2/5$ of the old. If we use the better DR algorithm, we will get even a better result.

4 Example of DIS Visual Simulation Experiment with MMO

In DIS, nodes communicate with each other through the P2P structure, whether information that is useful is decided by receiver. If the information has no value for receiver itself, it will be deleted, so it may waste bandwidth of information transmission network. With the increasing of simulation node, communication data will increase exponentially. This data transmission occupies limited network bandwidth, so the real time of communication interaction data can't be effectively assured. Based on this fault, the above mentioned visual field partition interaction range method was used in examples, when the node needing communication with others, sender confirms interaction objects according to its visual field range, excluding irrelevant entity node to interaction range through the game engine or 3D visual simulation engine mature view of field calculate function. The network flow load during DIS network data transmission process will reduce available. Figure 5 shows its flow diagram and specific realization project design.

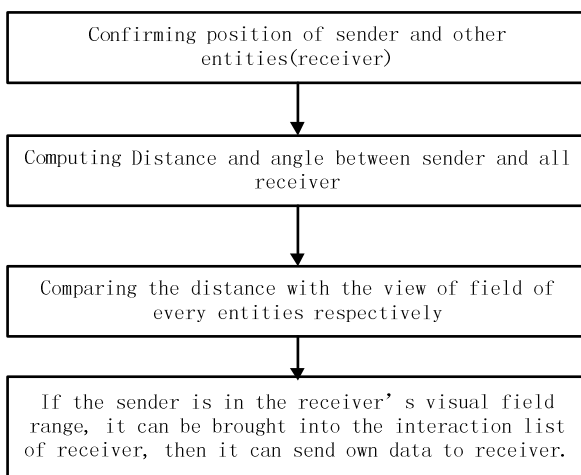


Fig. 5. Process of Interaction range of visual field partition

A visual simulation test scene was designed with 100 entities that are divided into 3 types: 5 helicopter, red, visual field is 8000m; 45 tanks, blue, visual field is 2500m; 50 Armored cars, green, visual field is 1800m. During DIS process, they need communication with each other. In order to getting convenience, in this research paper, we select the interaction data flow of the time t to conduct comparing study. Selecting node A (x, y) at will as the sender, send 1 time per second.

First: No matter how the visual field range is for each entity, entity A sends one data package to other 99 entities.

Second: A computes distance between itself with other 99 entities, if a distance is in the visual of field range of certain receiver, the receiver is selected as the communication target for A, then A can send data package to the selected entity. For example, the distance between A and B (B is helicopter) is 3000m, which is smaller than visual of field of helicopter (8000m), so A can send a state data package of itself to B during this network communication.

Figure 6 shows the distribution of 100 entities.

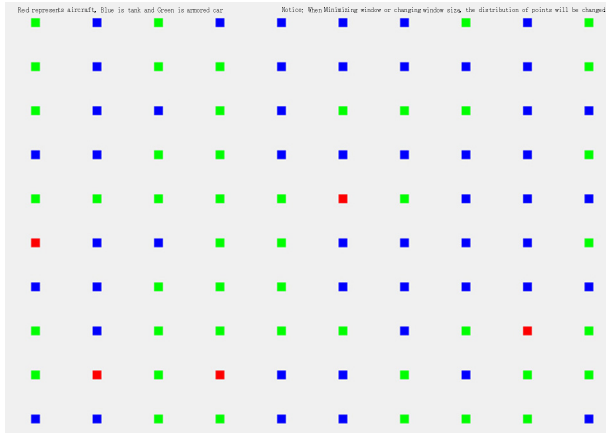


Fig. 6. 100 entities position distribution(1000m between two entities)

Table 1 gives the two results of the network data flow load which has been produced by the communication interaction.

According to above results, the network data flow load of first communication interaction experiment (that has no relation with entity node visual field range) is bigger than the second experiment (which use the visual field partition interaction range) approximately more than 20 times. In large scale visual simulation scene, with the increasing of entity quantity, the effect of reduce flow by using this solution will be more obvious.

Table 1. The load statistics of part network data flow in experiment

Coordinate	Entity Type	First (KB/S)	Second (KB/S)
(1, 1)	Armored car	0.8790	0.0401
(1, 2)	Armored car	0.9767	0.0596
(1, 3)	Armored car	0.9767	0.0596
(1, 4)	Tank	0.9767	0.0889
(1, 5)	Armored car	0.9767	0.0596
(1, 6)	Helicopter	0.9767	0.6846
(1, 7)	Tank	0.9767	0.0887
(1, 8)	Armored car	0.9767	0.0596
(1, 9)	Armored car	0.9767	0.0596
(1,10)	Tank	0.9767	0.0596

5 Conclusion

By studying the application of MMO mechanism in DIS visual simulation, we discuss the common method of DIS based on MMO mechanism and test its visual field partition interaction range technology by experiment. It has been applied in some large scale DIS visual simulation training system in which the author has participated, and received better effect.

References

1. Qing, X.: Equipment Combat Simulation Foundation. National Defense Industry Press, Beijing (2010)
2. Tang, S.Q.: The Study and Realization of Data Interaction Communicatio. In: DIS. Kunming University of Science and Technology (2005)
3. Sheng, G.Q., Wei, Z., Gong, Y.L.: DIS and Military Application. National Defense Industry Press (2003)
4. Wu, Y.-M., Zhou, W.-D.: Research on Distributed System Model of Massively Multiplayer Online Game. Computer Engineering 38(10) (2012)
5. Sheng, X.Y.: The Study of Key Technology in Large scale Network Game. Liaoning University Information Science and Technology College(110036)
6. Ou, L.B.: The Study and Realization Of Key Technology in Real Time Network Game. University of Electronic Science and Technology of China (2006)
7. Yi, C., Sheng, L.H., Ling, G.S., Cheng, L.B.: The DR Image Smooth algorithm Based on Bezier Curve. Microelectronics & Computer 25(5) (2005)

Towards the Representation of Virtual Gymnasia Based On Multi-dimensional Information Integration

Xiangzhong Xu¹, Jiandong Yang², Haohua Xu¹, and Yaxin Tan¹

¹ Simulation Center, Academy of Armored Force Engineering, No. 21 Dujiakan, Fengtai District, Beijing, China

² Department of Tracking and Control, China Satellite Maritime Tracking and Control Center, Jiangsu, China
xuxz02@21cn.com

Abstract. The multi-dimensional information integration of virtual gymnasia has not gained enough attention so far. It designs and implements the generation algorithms of space elements of the virtual gymnasium which efficiently solve the thorny issues, viz. the identification and the mutual map between the plan model and the stealth model of various space elements. A typical case is designed and implemented via SolidWorks, Microsoft Foundation Class (MFC), MultiGen Creator and OpenGVS based on multi-dimensional information integration where around 8000 seats in the very limited space have been designed and a seamless interaction between its plan model and stealth model is performed. Thus, users may not lose their orientations while navigating the virtual gymnasium. The research results can be applied to the on-line demonstration of sports venues, campuses and cities, etc., especially to the ticket-booking system.

Keywords: Multi-Dimensional Information Integration, Virtual Reality, Plan Model and Stealth Model.

1 Introduction

The Virtual reality abounds in awareness and has the characteristic of 3I,(Immersion, Interaction and Imagination). Thus, it gives rise to a vivid virtual world and has gained increasingly wide applications in various domains such as the three-dimensional visualization of space information [1], city plan [2], and the virtual battlefield environment [3]. The Virtual reality has brought brand new experiences to users.

Nowadays the three-dimensional modeling and representation of the virtual gymnasia have been researched profoundly in many literatures. However, the multi-dimensional information integration of virtual gymnasia has not gained enough attention so far.

The rest of this essay is organized in the following ways: Section 2 provides, in a brief form, the analysis of the building of the plan model for the virtual gymnasium which involves the design of the plan model by SolidWorks, the modeling of the space elements and the implementation of the plan model of the virtual gymnasium by

MFC; in Section 3, the stealth model of the virtual gymnasium (which can be roamed in several ways such as manual operation, automatic control and record and replay of the defined roaming path) is built by MultiGen Creator and managed by OpenGVS; in Section 4, the multi-dimensional information integration of the virtual gymnasium is performed based on the given technical framework. The users can experience the seamless interaction between their plan models and stealth models; finally the conclusion is found in Section 5.

2 The Build of the Plan Model of the Virtual Gymnasium

2.1 The Design of the Plan Model

The plan model basically reflects the shape, sizes, and the distribution of the seats and other space elements of the virtual gymnasium which lays a solid foundation to the stealth model of the virtual gymnasium. As a rule, the plan model should be correct, brief and clear.

It aims at the modeling of the football gymnasium. The virtual gymnasium is 195 meters long and 115 meters wide, and mainly consists of the court and the stand. The court is 80 meters long, 50 meters wide. The design capacity of the stand is 7984 seats and the east, west, south and north areas of the stand are respectively divided into 8 sub-areas such as No.1 south sub-area, No.2 south sub-area, and No.8 south sub-area, etc. There is one aisle, which is 1 meter wide, between the adjacent sub-areas. There are 20 rows of seats in each sub-area and there is a circular aisle (which is 2.5 meters wide) between the first ten seats and the last ten seats. The vertical range and the horizontal range between the adjacent seats are 0.5 meters and 1 meter respectively.

According to the above dimensional parameters, the plan model of the football gymnasium designed by SolidWorks is depicted in Fig. 1.

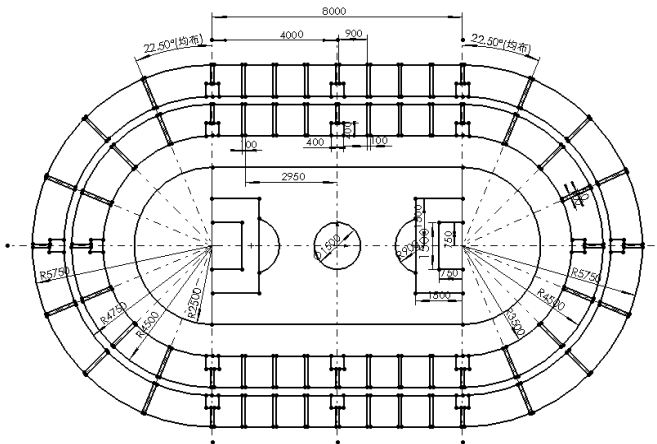


Fig. 1. The plan model of the football gymnasium designed by SolidWorks

Note that when the building the plan model, two measures are taken to consideration in order to ensure that every seat has as more pixels as possible: (1) the large football gymnasium is only one layer by design for simplicity where it is usually two layered; (2) the geometric sizes of the seats are augmented to its proper extent.

2.2 The Modeling of the Space Elements

The space elements of the football gymnasium, such as seats, aisles, circular runaway, court, stand, goals, midline and borders, should be respectively modeled as the identifiable objects. It focuses on the modeling of the stand, especially its seats.

The space elements of the football gymnasium such as seats, aisles, circular runaway, court, stand, goals, midline and borders should be respectively modeled as the identifiable objects. It focuses on the modeling of the stand, especially its seats.

2.2.1 The Model of Seats

Seats are the most important space elements for the ticket-booking systems; therefore the modelers usually pay the most attention their arrangements. The seats in the plan model and the stealth model must be mapped correctly; else many errors will be caused when performing the multi-dimensional information integration of virtual gymnasia. Thus, special attentions should be paid to the distribution model of the seats, the coding rules for the seats, the generation algorithms for the seats and the identification algorithms for the seats.

1 The distribution model of the seats

There are so many seats in the large virtual gymnasium that it will be a time-consuming, tedious and error-prone job to generate them manually. It is thus necessary to build the seats distribution model for the virtual gymnasium. A polar coordinate system can be used to represent the seats distribution considering the shape of the virtual gymnasium. It originates at the center point of the court as well as the horizontal and vertical positive axis point to the east and the north respectively.

2 The coding rules for the seats

To identify every seat, the coding rules for the seats are designed (Table 1). Each seat is assigned a unique code with 7 decimal bits and each seat code can be translated into the corresponding seat. There is a bidirectional map. The code is divided into 4 sections which stand for the number of the area, the sub-area, the row and the column respectively. For example, the seat code, 2051107, represents the seat which lies in the 11th row, the 7th column of the No.5 in the south sub-area.

Table 1. Coding rules for the seats

Codes(from left to right)	Bit 1	Bits 2-3	Bits 4-5	Bits 6-7
meanings	The number of the area	The number of the sub-area	The number of the row	The number of the column
remark	1-east,2-south,3-west,4-north	01-08 (clock wise)	01-20 (radiation)	01-08 (clock wise)

3 The generation algorithms for the seats

The north and the south sides of the virtual gymnasium are rectangle areas, and the other two sides of the virtual gymnasium are semi-circular areas. The generation algorithms for the seats in these two kinds of areas should be designed in different ways.

(1) The generation algorithms for the seats in the rectangle areas

To be brief, let's take the generation algorithms for the seats in the southern sub-areas as an example. As for the position in the southern sub-area with the coordinate (X_{n_0}, Y_{n_0}) , the center coordinate of the seat will be $P(X, Y, Z)$, and this transformation can be performed through the following formulae in group 1:

$$\begin{aligned} X &= X_{n_0} + L/2 + i * L \\ Y &= Y_{n_0} + m * W \\ Z &= m * H. \end{aligned} \quad (1)$$

Where n is the number of the southern sub-area ($0 < n < 9$) and

— m is the number of the row ($0 < m < 21$)

— i is the number of the column ($0 < i < 9$)

— L is the length of the stand (meter)

— W is the width of the stand (meter)

— H is the vertical range between the adjacent seats (meter)

(2) The generation algorithms for the seats in the semi-circular areas

The generation algorithms are a bit more complex. Again, in brief, let's take the generation algorithms for the seats in the eastern sub-areas as an example. As for the position in the eastern sub-area with the coordinate (X_{n_0}, Y_{n_0}) , the center coordinates with the seat will be $P(X, Y, Z)$, and this transformation can be performed through the following formulae in group 2:

$$\begin{aligned} X &= X_{n_0} + L/2 + \sin(\text{startA} + i * \text{endA}) \\ Y &= Y_{n_0} + r * \cos(\text{chairA} + i * \text{averA}) \\ Z &= m * H. \end{aligned} \quad (2)$$

Where n is the number of the eastern sub-area ($0 < n < 9$) and

— m is the number of the row ($0 < m < 21$)

— i is the number of the column ($0 < i < 9$)

— L is the length of the stand (meter)

— W is the width of the stand (meter)

— H is the vertical range between the adjacent seats (meter)

— r is the distance from the row to the center of the circle in the right respectively

— startA and endA are the starting angle and the ending angle of the column

— chairA is the angle of the row

— averA is the average angle of these seats

4 The identification algorithms for the seats

The purpose of the identification algorithms for the seats is to generate the code of the corresponding seat according to the specific screen position. In fact, they are the reverse algorithms of the generation algorithms for the seats. First, they transform the screen coordinate system into the polar coordinate system; second, they work out the numbers of the sub-area and row where the seat lies respectively; third, they work out the column where the seat is found; finally, the code of the corresponding seat is computed out.

2.2.2 The Model of Other Space Elements

Besides the massive seats, other space elements of the virtual gymnasium such as aisles, circular runaway, court, stand, goals, midline and borders also play important roles in the improvement of the users' experience and should be modeled respectively. That is, their distribution models, coding rules, generation algorithms, and identification algorithms should be designed and implemented. As they are similar to the counterpart of the seats, they are omitted here.

2.3 The Implementation of the Plan Model of the Virtual Gymnasium

The plan model of the virtual gymnasium is implemented on the platform of Microsoft Foundation Class (MFC) through the help of the Graphics Device Interfaces (GDI) as depicted in Figure 2. The key is to solve the problem of the transformation between the polar coordinate system and the screen coordinate system so as to "install" a great deal of seats in the space limited screen.

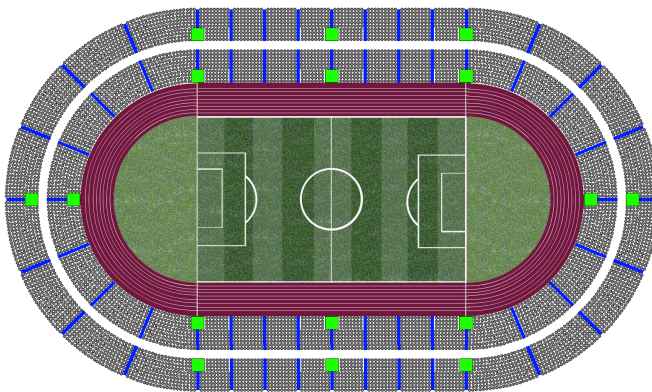


Fig. 2. The implemented plan model of the virtual gymnasium by MFC(a bird's-eye view)

In the above figure, all the seats, aisles and other space elements are modeled as objects and are all identifiable and can be positioned. Therefore, it lays a solid foundation for the seamless interaction between the plan model and the stealth model. Furthermore, it is helpful to other advanced applications such as ticket-book.

3 The Building of the Stealth Model of the Virtual Gymnasium

Nowadays, there are two mainstream approaches for the three-dimensional modeling [4]: Geometry-Based Modeling and Rendering (GBMR) and Image-Based Modeling and Rendering (IBMR). The great advantage of GBMR is that the users can directly watch the transformed screen pictures as the viewpoint changes; and its weakness is that it is difficult to build the models. At the same time, the great advantage of IBMR is that it is relatively easy to build the models; and its weakness is that the picture must be redrawn when the stealth changes.

3.1 The Implemented Stealth Model of the Virtual Gymnasium

The stealth model of the virtual gymnasium is built on IBMR on the platform of MultiGen Creator 2.5 based on the above plan models. A bird's-eye view of the virtual gymnasium is given in Figure 3.

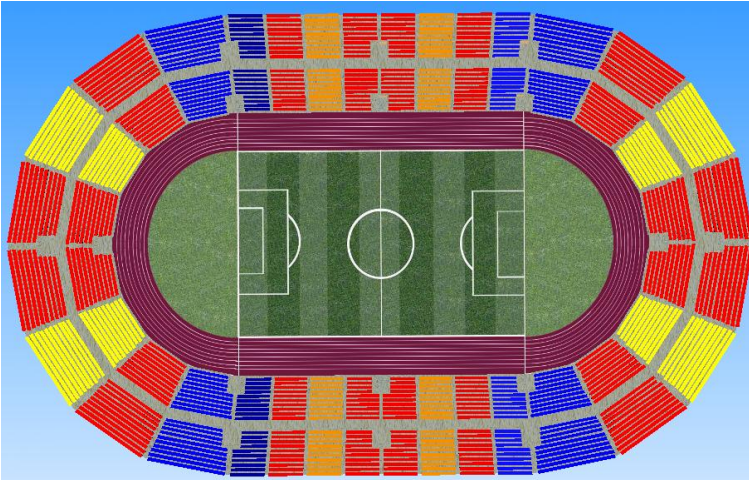
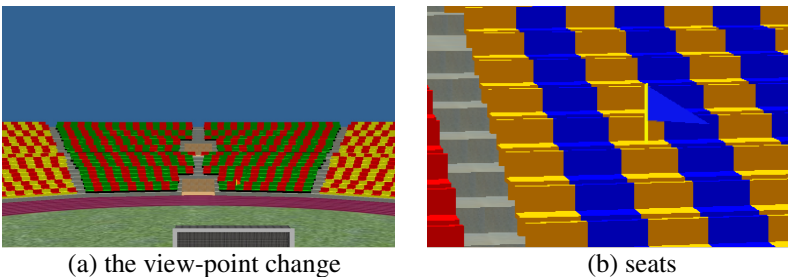


Fig. 3. The implemented stealth model of the virtual gymnasium by Multigen Creator (a bird's-eye view)

3.2 Roaming the Virtual Gymnasium

To fully demonstrate the usefulness of the virtual reality, the three-dimensional model of virtual gymnasium should be navigated. The three-dimensional model can be managed and presented to the users through the graphics engine such as OpenGVS. The users can roam the virtual gymnasium by changing the view-point through the arrow keys, including rotate, backward/forward and ascent/descent (Fig.4 (a)), or conduct themselves to a specific seat through the feature of the seats (Fig. 4(b)).



(a) the view-point change

(b) seats

Fig. 4. Roaming the virtual gymnasium

The users can roam the three-dimensional model of the virtual gymnasium such as walking through the virtual gymnasium, observing and experiencing it at different positions through the technology of “picture pick-up.” The nut is how to acquire the information about the mouse click on the seat visually in the virtual environment. So, the Screen-to-World algorithm has to be correctly implemented based on the basic principle of scenography projection.

The main functions of the implemented roaming system of the stealth model include are: (1) interactive roaming, that is, the users can control the roaming path through the mouse and the key by themselves; (2) automatic roaming, that is, the users can define the roaming path on the plan model, then the roaming system will walk through the stealth model according to the predefined path without human intervention; (3) the record and playback of the roaming path, that is, the roaming path used in the above-mentioned function will be recorded and can be played back for replay. It is obvious that the implementations of these functions rely heavily on the multi-dimensional information integration of virtual gymnasia which is of particular interest in this context.

4 The Multi-dimensional Information Integration of Virtual Gymnasia

The Plan View Display (PVD) and the stealthy view are the two most frequently used means for information display. The former will be helpful to build the global awareness of the virtual gymnasium and has low requirements for the hardware while it is difficult to represent the solid gymnasium vividly. The latter is good at representing the local awareness of the virtual gymnasium with restrictions to the limited computing speed and the screen size, thus, roaming is often essential for practical operations. Therefore, if the PVD, stealthy view, text, picture and sound can be seamlessly integrated, plus the interactive means, the multi-dimensional information space will be presented to the observers who will be inspired to acquire the knowledge and led to the creative thinking.

With the development of the technologies of the computer graphics, the three dimensional modeling, the rendering and the multimedia, it is possible to implement the multi-dimensional integrated information system [5] [6]. As for the virtual gymnasium, the principle of the interaction between the plan model and the stealth model can be described in Figure 5: the viewpoint in the stealth model will be automatically changed to the corresponding position pointed out on the plan model by the mouse, at the same time, the current position will be automatically displayed on the plan model accordingly when users select the specific space element on the stealth model. In other words, the two views are correlated to each other.

Note that the underlying two-dimensional and three-dimensional databases store all the required spatial and attribute data. They are the key to the multi-dimensional information integration of virtual gymnasia and can be constructed upon the basis of the spatial databases or the relational databases.

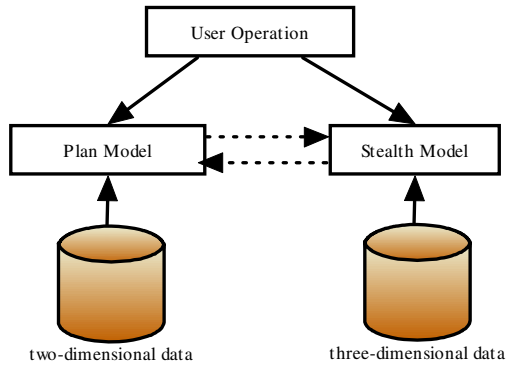


Fig. 5. The interaction between the plan model and the stealth model of the virtual gymnasium

5 Conclusions

So far, virtual reality has greatly affected the world and the way people live and work. With the help of various relevant technologies, the two mainstream information-representing approaches, the plan model and the stealth model of the virtual gymnasium can be built as demonstrated in this essay.

The PVD and the stealthy view are the two complementary not contending technologies. They should, and can, be seamlessly integrated to provide the users with a more complete and interactive view of the world for better understanding of the real world besides for the purpose of entertainment.

References

1. Yong, C.: The Three-dimensional Visualization of Space Information and Space Analysis Based on Enhanced Reality (in Chinese). *Journal of System Simulation* 19(9), 1991–1992 (2007)
2. Fang, X.: The three-dimensional simulation design system for city plan based on virtual reality (in Chinese). *Computer Simulation* 24(3), 230–234 (2007)
3. Chen, P., Wu, L., Yang, C.: The representation of the effecting range with the influence of terrain in the virtual battlefield environment (in Chinese). *Journal of System Simulation* 19(7), 1500–1503 (2007)
4. Chen, Y., Sun, Y., Niu, B.: On the three-dimensional stealth modeling and demonstration technologies based on the CAD data (in Chinese). *Journal of System Simulation* 19(7), 1504–1506 (2007)
5. Foley, J.D., Van Dam, A., Feiner, S.K., Hughes, J.F.: *Computer Graphics: Principles and Practice*, 2nd edn. Addison-Wesley, USA (1997)
6. Stefan, H., Rene, S.: From Flatland to Spaceland concepts for Interactive 3D Navigation in High Standard Atlases. In: 20th International Cartographic Conference. Science Press, Beijing (2001)

Night Vision Simulation of Drive Simulator Based on OpenGVS4.5

Zheng Changwei, Xue Qing, and Xu Wenchao

Academy of Armored Forces Engineering, Beijing, China
zhw_byh@163.com

Abstract. Night raid became more and more important in modern conflict because of the wide use of different kinds of night vision equipments; this is why the night training is playing an important role in military training. As an important method of military training, simulated training could not lose sight of night training. In this essay, we will consider the followings: firstly, the light properties of real time 3D graphics engine in OpenGVS4.5; secondly, the simulation of the vision effect of night environment, low-light-level night vision, and infrared night-vision scope; and finally, the results of our research that have been applied in a variety of driver training simulators and the satisfying effect of the training.

Keywords: Night vision simulation, low-light-level night vision simulation and infrared thermal image simulation.

1 Introduction

In recent years, the 3D graphic and simulation technology had been continually improved since different kinds of training simulator became more and more acceptable. With the improved fidelity of simulators, the 3D graphic and simulation technology became an efficient and important training method that has the characteristic of safety, environmental friendly, all-weather suitable and low cost. Driving training simulator offered a wide range of subject according to the requirements of driving: the whole process of driving is simulated, and all necessary skills for driving are focused on. The Driving simulation training had been an important auxiliary method of driving training, especially for tank driving training.

Night fighting is common in modern war, and is becoming more and more important. Night training is becoming an important training part in order to better prepare our army for the war in the future. In tank driving training, more than one-third is driving training at night. Therefore, night training simulation would directly affect the overall performance of the driving simulation.

2 OpenGVS4.5 Light Source Model

OpenGVS was one of the Quantum3D company's products which developed a real-time visual simulation scene graph. OpenGVS provided a variety of visual simulation functions

which were directly upon the three-dimensional graphics engine of OpenGL and Direct3D [1]. OpenGVS offered API based on its own resources, and can easily organize elements of visual simulation. Light source tools in OpenGVS can be used to control the dynamic lighting effect in the scene, which was established on the basis of the OpenGL lighting model. The purpose of the light source tool was to control the light in the scene to simulate the real-world lighting effects. The light source tool provided color, shading, and controlled the light in simulation scenario. The function of *GV_lsr_create* could create a light source. The function returned a handle that can be used in almost all other light utility functions to manipulate the particular source. The light source can be divided into an infinite light, local light, spotlight and point light according to the characteristics [2].

2.1 Infinite Light

All the default light sources were infinite light (also called directional light), the concept was from the light source of infinity "light." Since the light source was infinity far from an object, the light was considered to be parallel. That was very similar to the real world of sunlight. The light source can be set up by the following statement:

```
G_Vector sun_direction = {1.0, 0.0, 0.0}; // Light direction
GV_Light sun;
GV_lsr_create( &sun ); // Create an infinite light
GV_lsr_set_name( sun, "SUN" ); // Setting Name
GV_lsr_set_direction( sun, &sun_direction ); // Set direction
```

The infinite light sources could simulate real world when the sun at different times from sunrise to the downhill, as shown in figure 1.



Fig. 1. Simulate different time with unlimited light source

2.2 Local Light

The local light (also known position light) requires precise three-dimensional world coordinate position, the local light had a local effect depends on the relative strength and position of the light source. The local lighting effects depended on the location of the local light and the object and the scope of the local light.

```
G_Position light_position = {0.0, 5.0, 0.0};
G_Vector light_direction = {1.0, 0.0, 0.0};
GV_Light room_light;
```

```

GV_lsr_create( &room_light ); // Create a local light
GV_lsr_set_name( room_light, "ROOM " );
GV_lsr_set_position( room_light, &light_position ); // Set Position
GV_lsr_set_direction(room_light, & light_direction );// Set direction

```

2.3 Spotlight

One can also set the light to make it look like the spotlight. For example, set the shape of the light source a conical shape. The spotlight was actually a point light source with irradiation direction and coverage. It's the same to create an infinite light, create a spotlight needs specific property setting functions.

2.3.1 Irradiation Direction of the Spotlight

The irradiation direction of the spotlight was different from infinite light source. Direction of the light spotlight means the direction of the light.

```

G_Vector3 light_direction = { 0.0, 0.0, -1.0 };
GV_lsr_set_spot_direction(spotlight, &light_direction );

```

The above procedure sets the irradiation direction of a point source from the point light source to Z-axis negative direction.

2.3.2 Position of Spotlight

The light source position was the position where the light source was, the meaning of which was the same with the local light position.

```

G_Position world_position = { 0.0, 0.0, 0.0 };
GV_lsr_set_position(spotlight, &world_position );

```

2.3.3 Cone Angle of the Spotlight

By default, the angle of a point light source is 180 °, light launch in 360 °, so it was not a cone. It is necessary to specify the angle between the sides of the cone along the axis of the cone.

```

float cutoff_angle = 30.0 * G_DEG_TO_RAD;
GV_lsr_set_spot_cutoff(spotlight, cutoff_angle );

```

2.3.4 Attenuation Parameters Spotlight

There are four parameters of light attenuation, attenuation constant, linear attenuation parameters, quadratic attenuation parameters and light intensity index [1]. The spotlight light attenuation parameters were from scene after commissioning. The attenuation constant, linear attenuation parameters and quadratic attenuation parameters directly affected the size of the parameter of the light intensity index, and further affected the intensity of the light source [2].

```

float d = 50.0; // the limit distance of light irradiation
float kc = 0.5; // attenuation constant
float kl = 0.0001; // Linear attenuation parameters
float kq = 0.0001; // quadratic attenuation parameters
float attenuation = 1.0 / (kc + kl*d + kq*d*d);

```

```

GV_lsr_set_attenuation( spotlight , GL_CONSTANT_ATTENUATION, kc );
GV_lsr_set_attenuation( spotlight , GL_LINEAR_ATTENUATION, kl );
GV_lsr_set_attenuation( spotlight , GL_QUADRATIC_ATTENUATION, kq );
GV_lsr_set_spot_exponent( spotlight , attenuation );
2.4 Pixel-level point light

```

Spotlight is a good model for analog lights and headlight. While the series of OpenGVS real-time visual model-driven engine spotlight had bugs, such as rendering by the presence of surface and limited number. For this purpose, OpenGVS4.5 provided a Pixel-level point light. The model had the same properties with spotlights model which also allowed the user to set the different positions as required, the direction, the taper angle, the degree of attenuation and the irradiation distance in order to generate the pixel-rendering of lighting effects, and allowed the user to modify in real time according to the vehicle motion state. However, pixel-level point source model requires hardware support. Created as follows:

```

GVU_Plsr spotlight = NULL;
G_Boolean plsr_support ;
GVU_plsr_inq_supported( &plsr_support ) ; // Asked hardware support
if (!plsr_support) return NULL;
GVU_plsr_create( &spotlight);

```

3 Night Vision Effects simulation

Night training simulation needs the night vision of simulator. It's easy in the OpenGVS4.5 because the infinity light source in OpenGVS4.5 is very similar to the real world of the sun - as long as the proper direction vector set of infinity light can simulate dawn, sunrise, noon , evening, midnight, etc. We would like to indicate that in order to achieve these effects, the illumination attribute of all terrains, features, and the entity model must be turned on.

4 Night Observation Effect Simulation

Another problem that needed to be solved was the simulation of a variety of observation equipment in night training simulation. The existing observation equipment varied, the simulation methods were not the same with different principles and observation effect.

4.1 Headlight Illumination Simulation

Headlight illumination simulation could use OpenGVS4.5 unique pixel-level point source.

4.1.1 Creation and Initialization of Pixel-Level Point Light Sources

As mentioned earlier, as long as the hardware supports (currently this is not a problem, most of the high-end graphics cards are supported), one can use the pixel-level point source to simulate lighting effects. Properties of pixel-level point source model was the

same with spotlight as long as the appropriate setting of the irradiation direction, the light spot position, the cone angle, the attenuation constant, linear attenuation parameter, the second parameter and the irradiation distance attenuation parameters [2].

4.1.2 Synchrotron Light Source and its Subsidiaries Object

With tank lights, for example, to synchronize the point light source and the tank, which means the point source moved with the tanks. The process could be divided into two parts: (1) the position of the point source is synchronized with the tank, i.e. a point source is always maintained at the front side of the tank; (2) the rotation of the light irradiation source direction is synchronized with the tank and the pitch synchronization.

From the mathematical perspective, to ensure the synchronization a complex mathematical model was required in order to coordinate transformation. In fact, the DOF (Degrees of Freedom) node can easily achieve the synchronization. When construct a tank model, one needs to set the tank lights a DOF node, and set on the corresponding coordinate origin and axis orientation. The function of `GV_obi_inq_by_name_relative ()` in OpenGVS can find the node handle of DOF headlight, then use the function of `GV_obi_inq_pos_rot_world ()` can find the position coordinates and angle of each axis in world coordinate system, use the position coordinates and the axes corner set point source's pixel location coordinates and the illumination direction, you can easily achieve the synchronization.



Fig. 2. Tank headlights lighting simulation

4.2 Low-light-level Night Vision Simulation

Low-light-level night vision received faint light reflected from the object at night, and enhanced the faint optical image's brightness for the human eye to be observed [3]. In the night vision image, due to the role's image intensifiers, the objects in the scene are no longer showing the original color attributes, but showing different brightness green. The image resolution is low, only about 5% of visible light, and the instrument itself has a system noise on the eyepiece with white highlights blinking. [3]

We can simulate the low-light-level observation vision with the translucent green ambient light filters. In front of the viewpoint camera placed a bright green paste textured translucent filters and set the scene ambient light green could generate scenes with color filters green light in the visual texture color filters after fusion. And the random function can generate noise in night vision changing in the visual noise. The noise point in night vision could greatly improve the fidelity of simulation.

In addition, in order to ensure the correctness of the greatest visual distance, adding an appropriate concentration of fog in the scene model can limit the visual distance.



Fig. 3. The simulation of the effect of low-light-level night vision

4.3 Infrared Thermal Imaging Simulation

Thermal radiation was a common form for heat transfer; as long as the object was not absolute zero, and the heat radiation adapts to the temperature all the time. Thermal radiation was also electromagnetic wave, that invisible hotline was infrared rays [4]. Infrared Equipment converted the infrared thermal radiation to images that adult eye can recognize. Early infrared observation equipment had poor performance in the infrared image converter tube which made it difficult to identify with its own heat radiation. Therefore, an infrared searchlight on the viewing direction was needed; and these infrared observation equipment was called active infrared observation instrument. After the 1980s, the plus infrared radiation was no longer needed, and the identify ability was greatly enhanced; this infrared observation equipment were called passive infrared observation equipment (also known as thermal imager).

Because the thermal imager depended on the thermal radiation, its effect depends on the intensity of thermal radiation of various objects. The intensity of the radiation not only depended on the wavelength and temperature, but also has a close relation with the material properties of the object. Thermal Imager simulation is to establish a correspondence between different objects and imaging brightness. Research on theoretical modeling of IR had lasted several decades; the methods vary, but in general, can be divided into three types: firstly, statistical laws based on a very large number of species;

secondly, direct theoretical calculation based on the classical equations of heat transfer; and thirdly, a simplified theoretical calculation method called semi- experience, semi-theoretical method. From the view of real-time visual simulation, the simulation means to display the effect of variety of materials. The analysis of each graphic elements created in scene (terrain, buildings, vegetation and artificial entities, etc.), according to their characteristics in the infrared image conditions. When modeling in MultiGen Creator, one needs to appropriately set the IR imaging attribute (this is a gray- chromatography 4096); when OpenGVS imported the model, the infrared properties of the model needs to be opened, one needs to set the infrared attribute on, then can get the thermal imaging results. In the process of this, a lot of modeling techniques are involved under non- real-time status, not tired given the space. For analog active infrared observation device, one can add the green pixel point light source to simulate infrared searchlight illuminated .



Fig. 4. Active infrared simulation results

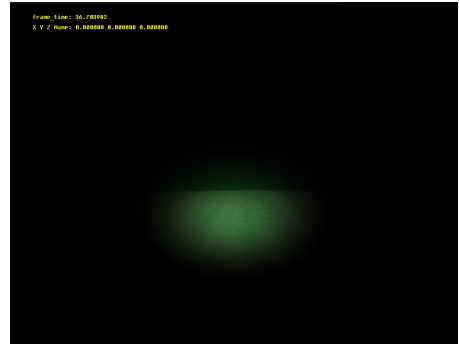


Fig. 5. Passive infrared simulation results

5 Conclusion

The results of the night driving training simulation shown in this research paper have been applied in a variety of driver training simulators, and received very good training effect. The results can also be applied to other training systems (e.g. fire training simulators) night vision visual simulation.

References

1. Quantum3D, Inc., OpenGVS 4.5 Programming Guide, Quantum3D, Inc. (2001)
2. Liu, G.: Based on Research OpenGVS4.3 of Lighting Effects, Armored Force Engineering Institute Graduation Thesis (2003)
3. Wang, J.: Night Vision Simulation of Armored Force Engineering Institute Graduation Thesis (2002)
4. Zhang, Y.: Simulation of Infrared Imaging Armored Force Engineering Institute Graduation Thesis (2002)

The Design and Implementation of Military Plotting System Based on Speech Recognition Technology

Wei Shao, Guanghui Li, Xiyang Huang, Qiang Liang, and Hesong Lu

Academy of Armored Force Engineering, Beijing 100072, China
v4rfvbhu8@163.com

Abstract. According to the application demand of military plotting, in this paper we design the intelligent military plotting system using speech recognition technology and MGIS, the military personnel can plot using speech. We analyze the function and performance of the mainstream speech recognition software and select the Viavoice, then we design the construction of intelligent plotting system, we mainly focus on the design of plotting and edition command, introduce the preparation work of the intelligent plotting system including data preparation and the setting of Viavoice, at last we realize the system using MGIS and Viavoice.

Keywords: Speech recognition, Viavoice, military plotting, intelligence.

1 Introduction

In the military plotting work, we can use military symbol and note to plot military situation on topographic maps, aerial photographs and maps [1]. With the development of information technology, by using computer plotting, we can effectively improve the efficiency of plotting, computer plotting is widely used in military field.

With the development of human-computer interaction technology, at present a new human-computer interaction method has appeared, it is speech recognition. By using the technique, the computer can understand what people say, the computer can analyze the speech, and the computer can implement corresponding command according to the speech. The technology has developed quickly, it has been used in the computer, television, mobile phone and other intelligent equipments. If we can combine speech interaction technology with computer plotting technology, then we can make the computer understand plotting command from plotting personnel, make the computer plot automatically. By using this technology, we can effectively reduce the workload of plotting personnel. From the considerations, we try to use the mature speech recognition software and military geographic information system(MGIS) to construct the intelligent speech plotting system, realize the automatic plotting.

2 The Analysis of Speech Recognition Software

If we want to construct speech plotting system, first we should confirm which speech recognition software is used, we should focus on the recognition rate and the difficulty of the speech recognition software. At present, the mainstream speech recognition software comprises Nuance, Viavoice, SAPI of Microsoft and the open source software HTK, these types of software are continuous speech recognition system for independent speaker, they have been widely used in the world. Some domestic speech recognition system is realized based on these types of software [2] [3] [4].

The SAPI of Microsoft is used commonly, at present the operating system Microsoft released comprises the speech recognition function, it can realize the text-speech conversion, but the recognition rate of the speech recognition software needs to be improved; Nuance is currently widely used in law and medicine, but it is difficult to develop, the advantage of the software is the high recognition rate [5]; HTK was developed by the University of Cambridge, the advantages of this software is open source, but it is also difficult to develop; Viavoice is developed by the IBM, the software has a high recognition rate, at present the software has been used in PDA and the intelligent vehicles, the software provides the interface of development, it can be embedded into a system, and it is not difficult to develop.

In order to construct the intelligent speech plotting system, we should select the mature speech recognition software. From the consideration, we choose the IBM Viavoice as the underlying platform for speech recognition.

3 The Structure of the Intelligent Plotting System

The basic structure of intelligent speech plotting system is shown below.

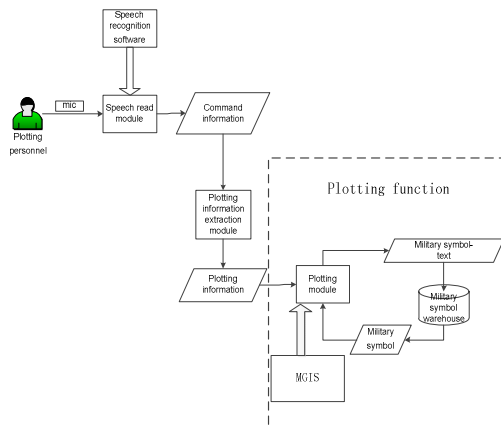


Fig. 1. The structure of intelligent plotting system

Speech read module is based on the speech recognition software, from the voice input equipment we can read the speech information of plotting personnel, then we can convert the speech information to the command text; the plotting information extraction module can draw the plotting information from the command text for the implementation of plotting; the plotting function is used to realize plot, we can use military geographic information system as platform, after plotting module obtains the plotting information, we can find the corresponding military symbol code from the data table, at last we realize the automatic plotting on electronic map using the military symbol.

4 The Design of Speech Plotting Command

4.1 The Element of Plotting Command

In accordance with the requirements of military plotting, if you want to use the military plotting, you need to determine the following elements: the color, size, location, direction and note of military symbol [6]. As long as the elements are determined, you can implement plotting.

The color of military symbol is determined by participating part sign, as the majority of military symbol denote our force are red, blue represents the enemy; we can determine the size of military symbol according to unit level; positioning is determined by the current position coordinate.

4.2 The Design of Plotting Command

From the analysis of plotting command element, in order to realize plotting, the plotting command should cover all the elements. According to the number of the positioning points, military symbol can be divided into punctate military symbol and linear military symbol, punctate military symbol only need one location point, linear military symbol needs two or more points, so we should design different plotting commands.

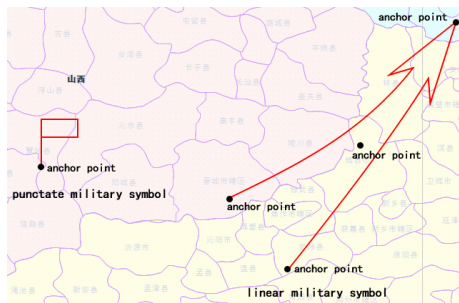


Fig. 2. The sketch map of different military symbols

According to the characteristic of punctate military symbol and linear military symbol, we design the plotting command as follow.

Table 1. The plotting command of punctate military symbol

Command name	Command content	Command criterion	Design
The plotting command of punctate military symbol	<p>plot 【battle part】 【military symbol name】 , location 【placename】 / 【longitude XXX, latitude XXX】 / 【y-axis XXX, x-axis XXX】 , direction 【the placename where the military symbol points to】 , note is 【military symbol note】 , execute.</p> <p>For example: plot red tank company, location y-axis is 3234957, x-axis is 16714370, direction is fengtai, note is tank B Company, execute.</p>	The command must include battle part, military symbol name and location.	<p>Punctate military symbol has one anchor point.</p> <p>【battle part】 : red or blue, red military symbol is red, blue military symbol is blue.</p> <p>【military symbol name】 : according to military symbol—name table, we can find the military symbol using military symbol name.</p> <p>【the placename where the military symbol points to】 : the placename where the mission points to.</p> <p>【military symbol note】 : it can be converted according to military symbol criterion.</p>

Table 2. The plotting command of linear military symbol

Command name	Command content	Command criterion	Design
The plotting command of linear military symbol	<p>plot 【battle part】 【military symbol name】 , anchor point 【number】 , 【placename 1】 / 【longitude XXX1, latitude XXX1】 / 【y-axis XXX1, x-axis XXX1】 , 【placename 2】 / 【longitude XXX2, latitude XXX2】 / 【y-axis XXX2, x-axis XXX2】 ... , note is 【military symbol note】 , execute.</p> <p>For example: plot red mechanized battalion, the number of anchor point is 3, y-axis is 3234957, x-axis is 16714370, y-axis is 3234967, x-axis is 16714388, y-axis is 3234947, x-axis is 16714358, note is tank A battalion, execute.</p>	【military symbol note】 need not be appointed.	<p>Linear military symbol has two or more anchor points.</p> <p>【battle part】 : red or blue, red military symbol is red, blue military symbol is blue.</p> <p>【military symbol name】 according to military symbol—name table, we can find the military symbol using military symbol name.</p> <p>【military symbol note】 : it can be converted according to military symbol criterion.</p>

According to the above two tables, we can extract the related elements of plotting from the speech command, and realize automatic plotting.

4.3 The Design of Edit Command

After plotting, sometimes we need to modify and delete the military symbol, therefore we need to set up the relevant edit command.

Table 3. Edit command

Command name	Command content	Command criterion	Design
Edit command	Selection: select 【battle part】 【military symbol note】 , execute. For example: select red tank A company, execute.	Similar command: select 【battle part】 【military symbol note】 , execute. Other types of command are useless.	The military symbol of selected unit can be placed in the center of the screen, and the military symbol is selected.
	Delete: delete 【battle part】 【military symbol note】 , execute. For example: delete red tank A company, execute.		According to 【battle part】 【military symbol note】 , delete appointed military symbol.
	Modify size: 【battle part】 【military symbol note】 , 【magnify】 / 【dwindle】 , execute. For example: red tank A company, magnify, execute.		When executing the command, the size of the military symbol is enlarged 2 times or reduced by half.
	Modify location: 【battle part】 【military symbol note】 , move to 【placename】 / 【longitude XXX, latitude XXX】 / 【y-axis XXX, x-axis XXX】 , execute. For example: red tank A company, move to y-axis 3234957, x-axis 16714370, execute.	Other types of command are useless.	When the military symbol is moving, the final location is the location of the first reference point of the military symbol, the size of the military symbol is unchangeable.
	Modify name: 【battle part】 【military symbol note】 , is named 【new military symbol note】 , execute. For example: red tank A company is named as tank B company, execute.	Other types of command are useless.	According to speech demand, modify military symbol note.

5 The Preparatory Work of Intelligent Plotting System

5.1 Military Symbol Data Preparation

From the analysis and the demand of plotting, we need to determine the military symbol, we can use a unique identification code to identify the military symbol, using this code we can prevent the military symbol from ambiguity problems.

If we use the digital coding mode to mark military symbol, such as the identification code of military symbol for a command post is 20001, then plotting personnel should remember 20001, the problem is the number is not easy to memorize, we should adopt a method which plotting personnel is familiar with.

According to the “The military symbol Regulations”, each military symbol has a name, and the name used in the military is common, so we should use the name of the military symbol, it accords with military personnel’s habit. Therefore, it is reasonable to use military symbol’s name in the plotting command. Combined with the need of computer recognition and people read, we can design the military symbol—name table in order to realize the intelligent selection of military symbol, the structure of the table is shown in the following table.

Table 4. The structure of military symbol - name corresponding table

Name	Type
Military symbol name	string
Military symbol library ID	int
Military symbol code	int

After the establishment of the corresponding table, we can use the plotting module to plot, because we can use the military symbol name to query the military symbol code.

5.2 Geographic Spatial Data Table

In the speech plotting command, there are three ways to determine the military symbol position, including latitude and longitude, Gauss coordinates and place names. By using the latitude and longitude coordinates and Gauss coordinates, GIS can navigate directly to the location, by using the place names, you need to find the place’s corresponding position coordinate.

In order to determine the corresponding position of the place, we should construct a corresponding table, this table is used to record the corresponding coordinate of places. The structure of the table is shown as follows.

Table 5. The corresponding table of place names

Name	Type
placename	string
longitude	float

Table 5. (Continued)

latitude	float
Gauss X	Long int
Gauss Y	Long int

Through the table, you can query the position coordinates of the place from the table.

5.3 The Preparation of Viavoice System

After we install the Viavoice software, we need to set the environment, volume and speech.

Through the voice input device, Viavoice can detect the noise of current environment, then recognize the speech. From the experiment, the Viavoice has high recognition rate in quiet environment.

Viavoice also can carry out the training. The users can read paragraphs provided by system, so the system can record and adapt to the users' accent which is helpful to improve the speech recognition rate [7] [8].

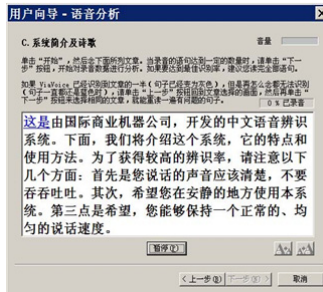


Fig. 3. The training of accent

After carry out the settings and training, we need to set the system as "dictation to the current application" which will provide the recognition text for plotting system, then the plotting system can deal with the text and realize automatic plotting.

6 System Implementation

6.1 The Development of MGIS

The underlying platform of the intelligent plotting is a certain type of military geographic information system, the development platform is based on VC++. The military geographic information system already contains map display, map operation (roam, zoom, etc), terrain analysis (distance, area), plotting, military symbol library and other related functions, and provides the military plotting, military symbol edit, military symbol delete and other development functions, user can achieve the required development.

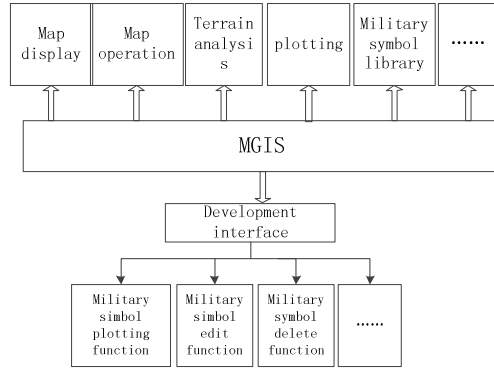


Fig. 4. The structure of MGIS

The main function of the intelligent plotting system includes [9]:

```
1 void CommandGet(char* m_commandtext )
```

The function is used to acquire the text information from the Viavoice.

```
2 void TextCope(char* m_getcommandtext )
```

The function is used to analyse the text, and call different functions according to the text content in order to realize the automatic implementation of the corresponding operation.

3 Plotting, delete function introduction

```
Char *MgsDrawDot (int libID, int code, FPOINT PT);
```

// We can use the function to plot the military symbol on the specified coordinates(only one specified location). The combination code libID and code uniquely identify the military symbol, pt is the latitude and longitude, the return of the function is only ID of military symbol.

```
Char *MgsDrawLine (int code, FPOINT *pt, int node-Count);
```

// We can use the function to plot the military symbol on the specified coordinates (two or more locations). Code is military symbol code, it is used to uniquely determine the military symbol, pt is the latitude and longitude array, nodeCount is the point number, the return of the function is only ID for military symbol.

```
int MgsDelObject (char* JbID);
```

// The function is used to delete the military symbol from the map which has the specified ID. JbID is the only ID of the military symbol.

If we extract the unit, coordinate and the military symbol from text information, then we can achieve automatic plotting using the functions above.

6.2 Implementation Effect

We use the military geographic information system and Viavoice to realize the intelligent plotting system, the effect is shown below.



Fig. 5. The fieldwork

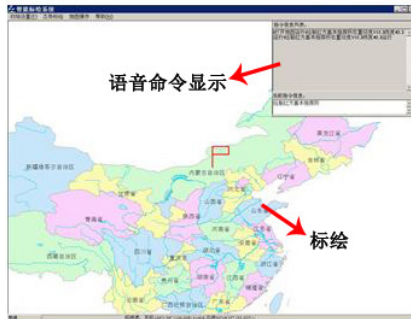


Fig. 6. The plotting effect

Intelligent plotting system can use computer and Mike equipment to realize the automatic plotting, judging from the experiment effect, if the plotting personnel pronounce correctly, the recognition rate will be higher, the recognition rate is also influenced by the environment noise. From the plotting effect, as long as the command covers the all elements, the location, shape and size of the military symbol will be reasonable.

7 Conclusions

Intelligent speech plotting system is focused on the correct rate, speech recognition rate, if the speech recognition rate is low, the recognition speed is low, the plotting system is not useful, the voice command also need to be simple and to meet military specifications. The military personnel want the system to be convenient and fast, if you design commands considering the technical demand blindly, without considering the practical need, the practical value of the system is also very limited.

Implementation of intelligent speech plotting is a beneficial attempt, it can enhance the efficiency of military plotting, but elements of speech command is needed to extract from the military or combat plan, it will increase extraction workload, the direction of further study is to research intelligent recognition from military plan text, computer can understand the semantic set of the plan, and it can extract the related elements of plotting and plot, it will omit the step of speech, this will improve the intelligence of the plotting system.

References

1. Wang, M., Wang, X.: The contract tactical plotting. Shijiazhuang Army Command College (2), 1–3 (2000)
2. Hermansky, H., Morgan, N.: RASTA processing of speech. *IEEE Transactions on Speech and Audio Processing* 2(4), 100–120 (1994)
3. Burget, L., Motlcek, P., Grezl, F., et al.: Distributed speech recognition. *Radio Engineering Prague* 11(4), 36–39 (2002)
4. Ramirez, J., Gorriz, J.M., Segura, J.: Improved voice activity detection based on integrated bispectrum likelihood ratio tests for robust speech recognition. *Acoust. Soc. Am.* 121(5), 1034–1221 (2007)
5. Ferrans, J., Engelsma, J.: Software architectures for networked mobile speech applications. *Automatic Speech Recognition on Mobile Devices and over Communication Networks* 304 (2008)
6. Young, S., Evermann, G., Gales, M., et al.: *The HTK book*(for HTK version 3.4). Cambridge University Engineering Department (2006)
7. Ren, Z.: The study And Realization of speech interaction method based on the SAPI engine. *Luoyang Industrial University* (1), 19–25 (2005)
8. Baidu Encyclopedia.Nuance
9. Xia, S., Xia, Z.: *The guide of plotting*, vol. (12), pp. 54–57. Military Science Press (2004)

Defect Detection in Fabrics Using Local Binary Patterns

Pengfei Li, Xuan Lin, Junfeng Jing, and Lei Zhang

Xi'an Polytechnic University, College of Electronical and
Information, 710048 Xi'an, Shaanxi, China
{Li6208, linxuan0122}@163.com,
{413066458, 11795503}@qq.com

Abstract. To detect defects in fabrics more efficiently, easily and accurately, a method based on Local Binary Pattern (LBP) is proposed in this paper. The main purpose of this algorithm is to extract the feature value of fabric images. Firstly the feature of the whole defect-free fabric image is got with LBP algorithm. Then the image is divided into small detection windows, and the feature of each window can be obtained. Compare their similarity calculated by Chi-square function to get the threshold. Then process the defective images according to the same procedure. At last compare the similarity with the threshold to obtain defect regions. The defects are detected at the same time. Experimental results demonstrate that, LBP algorithm is effective in the area of detecting defects of fabrics.

Keywords: Local Binary Pattern, feature value, similarity, defect detection.

1 Introduction

The development of global fabrics is very rapid in recent years. Simultaneously textile quality requirements for people are increasing. Consequently the quality of textiles needs to be controlled strictly. Defects have a serious impact on the identification of quality levels for fabrics [1]. According to statistics, if there is a defect, it may cause the value of fabrics decreased by 45%-60% [2], [3]. Therefore, defect detection becomes an essential step during fabric quality assessment.

With the continuous development of digital image processing technology, many domestic and foreign scholarships have conducted extensive researches on defect detection. To sum up, all the defecting methods can be divided into statistics based, model based and spectrum based. Statistical methods are based on the gray properties of both the pixels and their neighborhoods. Defects can be detected through studying the statistical characteristic in the texture area and determining the difference between the defect region and the normal region. I-Shou TSai et al. [4] have proposed a method for fabric defect detection using GLCM. With this method, threshold is not needed to determine the defect area of the detected fabric image, but a very large amount of calculation would exist. Model-based approaches describe the texture characteristics of fabrics by the parameters of particular models. B. S. Manjunath [5] has conducted a

study on fabric defect detection with the help of Gauss - Markov random field (GMRF) model. This method is not restricted by the types of defects, however, shortcomings also exist. The amount of calculating the estimated model parameters is very large and time-consuming is quite long. The methods based on spectrum are suitable for particular textured fabric. Wavelet transform [6], [7], [8], Gabor filters [9] as well as the discrete Fourier transform (DFT) [10], [11] are typical spectrum based methods of fabric defect detection. With such methods, the defect parts of the fabric image are highlighted mainly by time-frequency analysis and detected through the next thresholding. Nevertheless the time-frequency transform operations will reduce the detection speed, and has a large amount of calculation. As a result, it is difficult to apply such methods to detection online directly.

A method based on *LBP* is proposed in this paper. The global and local feature values of fabric images are extracted by *LBP* and further analyzed. Then the similarity can be calculated with Chi-square function. Set a threshold to determine the defect area. Thus, the purpose of defect detection is achieved.

2 Local Binary Patterns

LBP is originally put forward by Ojala [12] in 1999. For the reasons that *LBP* can be regarded as an effective texture description operator and its prominent ability of describing the local texture features of the images, *LBP* has been widely used in the area of the description of image texture.

The basic idea of *LBP* algorithm is built on pixels. Comparing the gray values of the center pixel with its surrounding neighborhood pixels, relative gray can be obtained and considered as the response of the center pixel. Therefore, *LBP* is invariable for monotonic gray-scale changes. Basic *LBP* operator is shown in Figure 1.

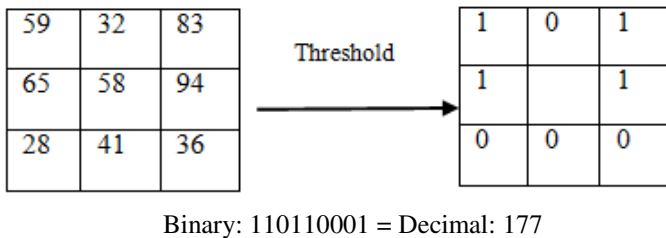


Fig. 1. Basic LBP operator

In this model, a neighborhood of the 3×3 window is regarded as the processing unit. The gray value of the center pixel is regarded as the threshold. The gray values of the surrounding eight neighbors pixels are compared with the threshold and processed by binarization. If the gray value of the pixel is smaller than the center pixel gray value, it

will be set as 0. Otherwise it will be set as 1. After the binarization process being completed, read the binaries obtained by thresholding from the top-left corner one by one clockwise. Then transform the string of binary into a decimal number, which is looked on as a response of the center point.

To meet the needs of different sizes of the texture feature information, Ojala et al. [13] have made a further improvement on the basic LBP operator. The original 3×3 neighborhoods can be extended to any neighborhoods, and circular neighborhoods are used instead of the previous square neighborhoods. As a result, any radiuses and numbers of neighborhood pixels could be obtained.

Fig. 2 shows three cases of circular LBP operator model, in which the radius R is 1, 2, and 3 respectively, and the number of neighborhoods P is 8, 16, and 24 accordingly. If the neighborhood point is just in the center of the grid, the gray value of the square in which the pixel located can be regarded as its own gray value directly. If the neighborhood point is located in a cross section of two squares, its gray value can be calculated by bilinear interpolation.

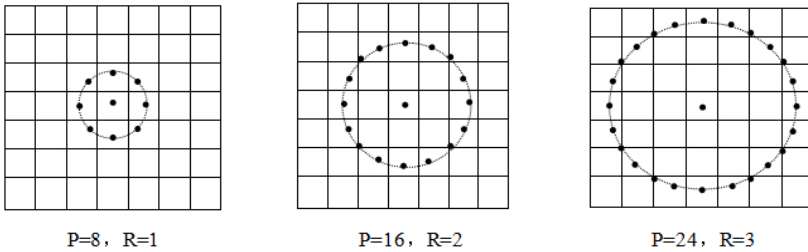


Fig. 2. R=1, 2, 3, P=8, 16, 24 circular LBP operator model

So since the general process of the LBP operator is as follows. Set an arbitrary pixel within a local area of an image is $f(x_c, y_c)$. And look on this pixel as the center pixel, whose gray value is g_c , and the gray values of the neighbors in the local unit are g_0, g_1, \dots, g_p respectively. The texture features T of the local region can be expressed as equation (1):

$$T \approx t(g_0 - g_c, g_1 - g_c, \dots, g_p - g_c) \tag{1}$$

Regard the gray value of the center pixel as the threshold, and the other pixels in the neighborhood unit are processed by binarization following the method defined as formula (2):

$$T \approx t(s(g_0 - g_c, g_1 - g_c, \dots, g_p - g_c)), s(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{2}$$

A P bits binary string is obtained after the process. Then convert the binary string into a decimal number. Thus, the feature value, namely the LBP value of the center pixel of a circular neighborhood whose radius is R and the number of neighborhood pixels is P can be got. The process can be expressed as Eq. 3:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \tag{3}$$

For the LBP operator model shown in Fig. 1, when the image is rotated, the arrangement of the data of the window will change accordingly. As a consequence, there are many cases of the LBP values obtained as the preceding steps occurring for a same image. To solve this problem, Tmeanpaa et. al^[14] have improved the original LBP algorithm, redefining the LBP operator as follows:

$$LBP_{P,R}^{ri} = \min\{ROR(LBP_{(P,R),i})\}, i = 0,1,\dots, P-1 \tag{4}$$

As seen from Eq. 4, after the image rotated for i times, select the minimum value of LBP as the neighborhood LBP value. Thus the improved LBP operator is invariable to rotation and can eliminate the impact during the image rotation.

According to Eq. 3, 2^P kinds of cases can occur after being dealt with the basic LBP algorithm to an image. To be specific, for $P=8$, $2^8 = 256$ kinds of cases will be got, and for $P=24$, the kinds of situations will reach to 2^{24} . So since, the dimension of LBP characteristic is quite high and the calculation amount is too large. It does not contribute to conduct follow-up studies. For this situation, Ojala^[13] has proposed the concept of uniform patterns. That is to say, for a local binary pattern string in its circulation state, if the times of the change between 0 and 1 are no more than twice, then this local binary pattern is regarded as a uniform pattern. Otherwise it belongs to a non-uniform pattern. It can be expressed as follows:

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c), U \leq 2 \\ P + 1, otherwise \end{cases} \tag{5}$$

Where,

$$U = |s(g_{p-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \tag{6}$$

From Eq. 5 and Eq. 6, the times of 1 appearing in a uniform pattern is considered as the LBP value of the neighborhood. For the non-uniform patterns, the LBP value of the neighborhood is set as $P + 1$. After improvements, the dimension of LBP values drops to $P + 2$. The dimension of the feature values is significantly reduced and the efficiency of the algorithm is effectively improved.

3 Defect Detection

The process of defect detection is divided into training phase and detection phase.

3.1 Training Phase

Select a defect-free image whose texture is as same as the sample detected. At first the entire defect-free image is disposed by *LBP* operator, and a $P + 2$ dimensional feature vector M can be got. In other words, M is the probability of $0-P+1$ occurring in the *LBP* value of the entire defect-free image. Then the image is divided into several non-overlapping windows of the size $W_d \times W_d$ pixel, which are called the detection windows. Moving step of detection windows is two pixels. The detection windows are processed with *LBP* operator (*LBP* mask). Thus, the feature vectors of each window can be obtained. In order to make the results more accurate, it should ensure that the calculated times of *LBP* mask in each detection window are not less than 100 times. Suppose the size of the *LBP* mask is $w_m \times w_m$, W_d , the size of the detection window should satisfy the following conditions:

$$(W_d - w_m + 1)^2 \geq 100 \quad (7)$$

LBP masks used in the experiments are $LBP_{8,1}$, $LBP_{16,2}$, $LBP_{24,3}$. The maximum of P is 24, accordingly the maximum of w_m is 7. As a result, W_d should be more than 16, namely $W_d \geq 16$.

For the next step, the similarity of S_k and M is calculated with Chi-square function. The formula is as follows:

$$L_k(S_k, M) = \sum_{i=0}^{P+1} \frac{(S_{ki} - M_i)^2}{S_{ki} + M_i}, \quad k = 1, 2, \dots, N \quad (8)$$

Where, S_{ki} represents the probability of i occurring in the *LBP* values of the k -th detection window. M_i represents the probability of i occurring in the *LBP* values of the whole image. N is the number of the detection windows. L_k describes the similarity of the k -th detection window and the whole image, and the smaller the L_k is, the more similar between the k -th detection window and the whole image. Select the maximum during the L_k as the threshold used to determine the defect windows:

$$T = \max(L_k), k = 1, 2, \dots, N \quad (9)$$

3.2 Detection Phase

The entire process of defect detection is expressed as follows:

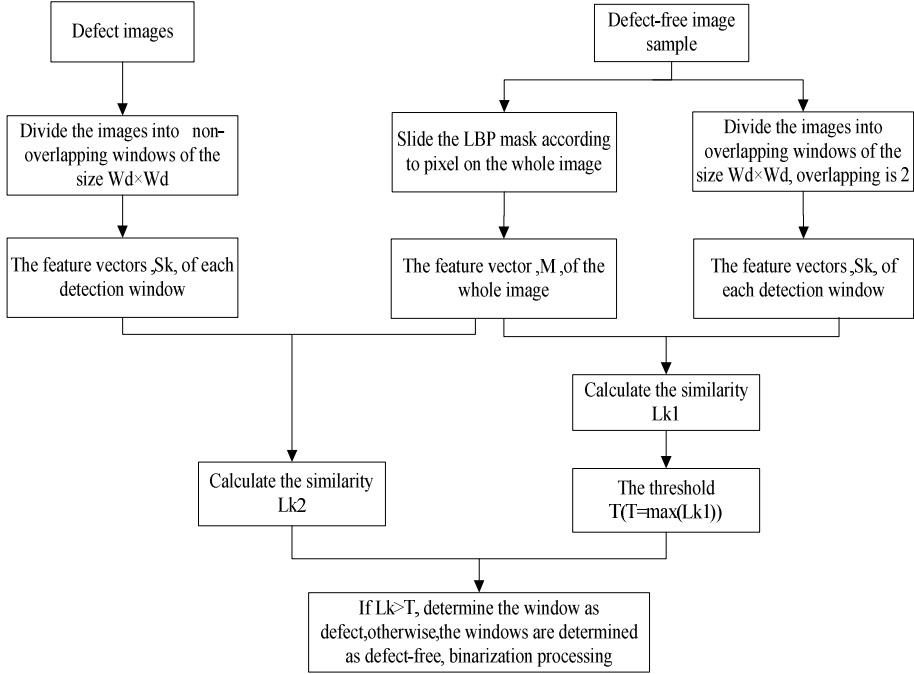


Fig. 3. Detection algorithm process

The defect image is divided into overlapping detection windows of size $W_d \times W_d$ pixel as the same method as processing defect-free image. The LBP mask is slid according to pixel in sequence for each detected window. The feature vector S_k of each detected window can be calculated at the same time. Calculate the similarity of each detection window and the entire image according to Eq.8. Compare L_k with the threshold T . If $L_k > T$, the detected window is judged as the defect window, in which all the pixels located are set as 255. Otherwise the detected window is judged as the defect free window, in which all the pixels located are set as 0.

4 The Experimental Results

The experiment is conducted with MATLAB 7.6.0. As experimental subjects to verify the enforceability of the proposed method, the test samples are selected from the TILDA database and Henry Y. T. Ngan of the Industrial Automation Research Laboratory of Hong Kong University. The size of each picture is 256×256 pixel.

Process the sample images with the steps of the training phase and the testing phase described above. $LBP_{8,1}$, $LBP_{16,2}$ and $LBP_{24,3}$ are respectively used to process the images to extract their feature values and detect the defects. In addition, to develop the accuracy of the detection, the overlapping step is 2. In the paper, ten kinds of backgrounds and defects images are analyzed as examples. The detection results are presented in Fig.4.

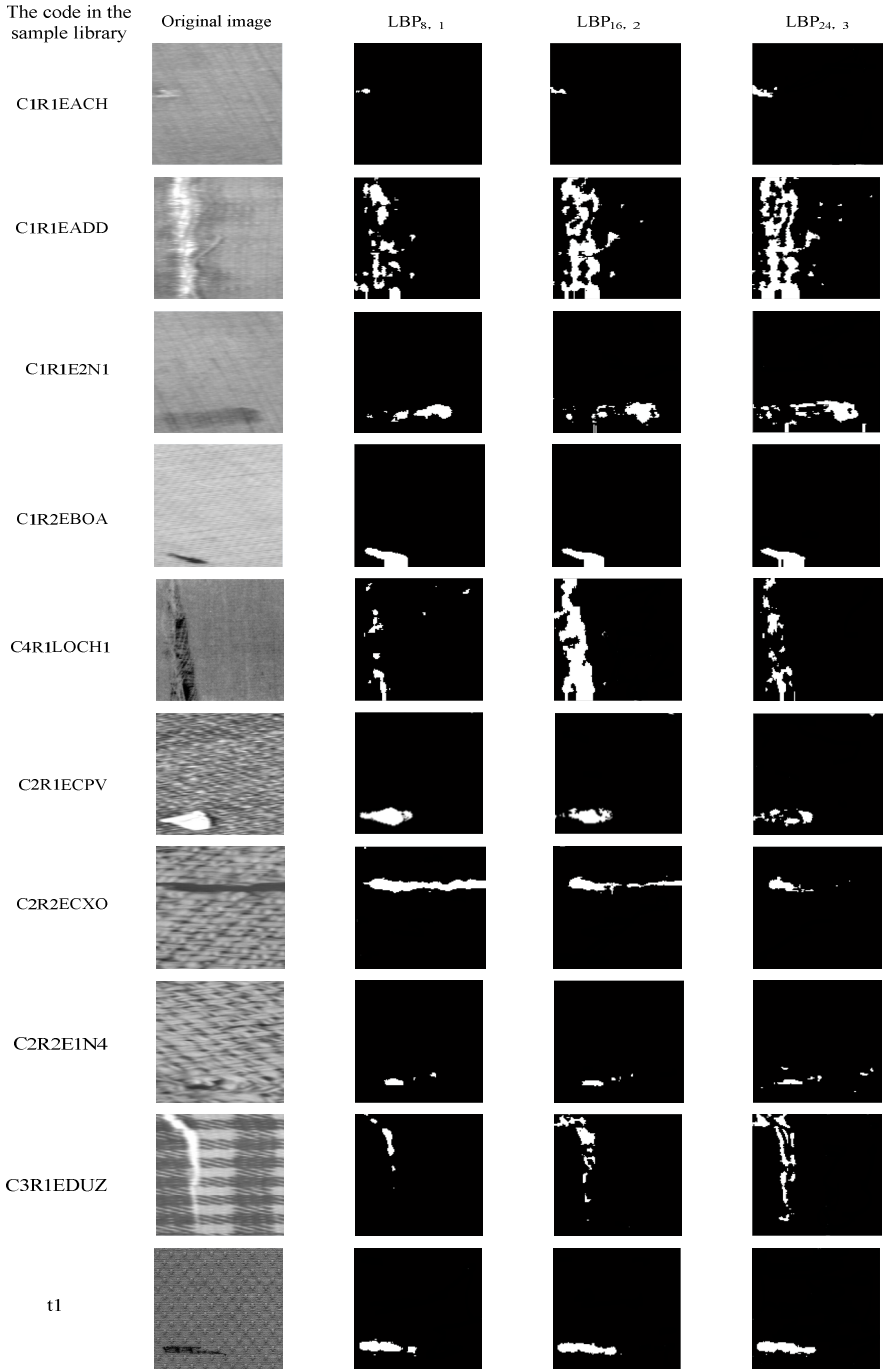


Fig. 4. The test results of some samples

To compare the detection accuracy of different parameters of the LBP, false detection rate is proposed as a measure scale to quantify the LBP operator. It is defined as follows:

$$E = \left(\frac{N_c + N_d}{N_{total}} \right) \times 100\% \quad (10)$$

Where, E represents the false detection rate. The smaller the value is, the better the detection algorithm does. N_c is the number of the windows which are detected as defects windows while they are defects-free windows. N_d means the number of the windows which are detected as defects-free windows while they are defects windows. N_{total} shows the total number of the windows which are got by dividing the defect image.

According to Eq.10, the false detection rate of the 10 images processed by $LBP_{8,1}$, $LBP_{16,2}$ and $LBP_{24,3}$ can separately be calculated, and the results are carried on an analysis, which are shown in Table 1.

Table 1. . The false detection rate of the 10 images processed by $LBP_{8,1}$, $LBP_{16,2}$ and $LBP_{24,3}$ separately (%)

Defect image	$LBP_{8,1}$	$LBP_{16,2}$	$LBP_{24,3}$	Average
C1R1E2CH (A)	4.21	3.89	3.17	3.76
C1R1EAD (B)	4.92	4.38	4.07	4.46
C1R1E2N1 (C)	5.28	4.52	4.06	4.62
C1R2EBOA (D)	4.83	4.27	4.42	4.48
C4R1CLOCH1	7.04	3.85	4.36	5.08
C2R1ECP (F)	2.89	4.74	5.97	4.53
C2R2ECXO (G)	3.32	4.73	6.54	4.86
C2R2E1N4 (H)	6.69	6.13	5.08	5.97
C3R2EDUZ (I)	5.74	4.86	3.51	4.70
t1 (J)	4.54	3.98	3.27	3.93
Average	4.95	4.54	4.45	

From Fig.4 and Table 1 it can be visually seen that the defect detection effect of $LBP_{24,3}$ operator is the most accurate. The average of false detection rate is 4.45%. When processing the defect images with $LBP_{16,2}$ operator, there are small parts of the leakage inspection windows existing by mistake. The average of false detection rate is 4.54%. While the average false detection rate is 4.95% when applying $LBP_{8,1}$ operator. The error detection section will be more than applying $LBP_{16,2}$ operator. However, with

the increase of P , the running time of the program is also increasing. In other words, $LBP_{24,3}$ consumes a longer time than $LBP_{16,2}$, which is a little slower than $LBP_{8,1}$. From the overall view, the detection results of employing LBP operator for the fabric of the delicate texture (e.g. defects A-E) are more obvious than for the fabric of the coarse texture. With the increasing of roughness of fabric texture, the testing effect will be the less desirable (e.g. defects F-H). In addition, the LBP algorithm is also available to the patterned fabrics (e.g. defect I-J).

5 Summary

LBP is proposed as a method to detect the fabric defects in this paper. The relative gray value between pixels is considered as response in LBP algorithm. Thus it is invariant towards the monotonous gray changing. The original LBP operator is improved to make it own rotation invariance and set the threshold to determine the uniform patterns. As a result, the algorithm can effectively reduce the dimensions of feature values and the computation time, because only the information of uniform patterns is analyzed. Make the similarity of the feature vectors of both the detection windows and the whole image as the criterion to confirm defects. In this way defects can be identified and segmented more accurately. Experimental results have shown that the detection results of LBP algorithm are reliable whether in the side of intuitive visual or in terms of the false detection rate. However, LBP algorithm is not available to detect all kinds of defects on all texture backgrounds. Consequently it still needs a further improvement in the area of fabric defect detection to accommodate different fabric texture backgrounds.

Acknowledgement. The authors gratefully thank the Scientific Research Program Funded by National Natural Science Foundation of China (61301276), Shaanxi Provincial Education Department (Program No. 2013JK1084), Shaanxi Science and Technology Research and Development Project (Project No.2013K07-32).

References

1. Li, L.Q., et al.: Image Processing Progress in Fabric Defect Automatic Detection. Donghua University (Natural Science), 李立轻等, 图像处理用于织物疵点自动检测的研究进展. 东华大学学报(自然科学版) 28(4), 118–122 (2002)
2. Cho, C.S., Chung, B.M., Park, M.J.: Development of real-time vision-based fabric inspection system. IEEE Transactions on Industrial Electronics 52(4), 1073–1079 (2005)
3. Kumar, A.: Computer-Vision-Based fabric defect Detection: A Survey. IEEE Transactions on Industrial Electronics 55(1), 348–363 (2008)
4. TSai, I.-S., Lin, C.-H., Lin, J.-J.: Applying an Artificial Neural Network to Pattern Recognition in Fabric Defects. Textile Research Journal 65(3), 123–130 (1995)
5. Manjunath, B.S., Chellappa, R.: Unsupervised texture segmentation using Markov Random Filed Models. IEEE Transactions on Pattern Analysis 13(5), 478–482 (1991)

6. Guan, S., Shi, X.: Fabric defect detection based on wavelet decomposition with one resolution level. C. In: International Symposium on Information Science and Engineering, Shanghai, pp. 281–285 (2008)
7. Sari-Sarraf, H., Goddard, J.S.: Vision system for on-loom fabric inspection. *IEEE Transactions on Industry Application* 35(6), 1252–1259 (1999)
8. Yang, X.Z., Pang, G.K.H., Yung, N.H.C.: Discriminative fabric defect detection using adaptive wavelet. *Optical Engineer* 41(12), 3116–3126 (2002)
9. Zhang, Y., Lu, Z., Li, J.: Fabric defect detection and classification using Gabor filters and Gaussian Mixture Model C. In: Zha, H., Taniguchi, R.-i., Maybank, S. (eds.) ACCV 2009, Part II. LNCS, vol. 5995, pp. 635–644. Springer, Heidelberg (2010)
10. Tsai, D.-M., Chao, S.-M.: An anisotropic diffusion-based defect detection for sputtered surfaces with inhomogeneous textures. *Image and Vision Computing* 23(3), 325–338 (2005)
11. Tsai, D.-M., Kuo, C.-C.: Defect detection in inhomogeneously textured sputtered surfaces using 3D Fourier image reconstruction. *Machine Vision and Applications* 18(6), 383–400 (2007)
12. Ojala, T., Pietikainen, M.: Unsupervised Texture Segmentation Using Feature Distributions 32, 477–486 (1999)
13. Ojala, T.: Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(7), 971–987 (2002)
14. Huang, F.: Face Recognition Based on LBP. D. 黄非非. 基于LBP的人脸识别研究. ChongQing University, ChongQing (2009)
15. Feng, X., Hadid, A.: Facial Expression Recognition with Local Binary Patterns and Linear Programming. *Patterns Recognition and Image Analysis* 15(2), 546–548 (2005)

Research on Teeth Positioning Based on Wavelet Transform and Edge Detection

Zhou Zhou, Guoxia Sun, and Tao Yang

School of Information science and Engineering, Shandong
University, Jinan, 250100, China
Zhouzhou0522@126.com

Abstract. Human Identification from Dental X-Ray Images is a new Biometric Identification technology based on the use of modern digital image processing technology. Tooth positioning is the initial step in the individual identification system using dental X-ray image, and its main purpose is to find the accurate position of tooth in the high-resolution X-ray images, which can reduce data redundancy and provide support for the establishment of the dental images database and subsequent processing. Automated positioning and cropping of dental X-ray records is a challenging problem due to the heterogeneity of dental records. This paper proposed an algorithm to process the dental X-ray images using wavelet transform and edge detection techniques, respectively in horizontal and vertical directions to find the position of tooth area. Simulation results proved the accuracy and effectiveness of the method.

Keywords: X-ray image, teeth positioning, wavelet transform, edge detection.

1 Introduction

Biometric recognition refers to identity authentication using human biological characteristics. Biological features usually have the qualities of uniqueness, measurability, heredity and invariance. Natural teeth have strong abilities of resistance to corrosion and degradation. The melting temperature of teeth is equal or greater than 1600 °C. Physical and chemical changes of the internal fine structure of teeth could not occur even when the teeth were burned near the melting point. This stability feature makes teeth becoming an important method to identify persons of disasters such as criminal cases and bombings [1].

Affected by modern medical image processing technology, dental images used for individual recognition are mainly X-ray images. And panoramic images are widely used for clinical diagnosis and treatments. In order to obtain the comprehensive diagnostic information, the range of the panoramic images generally includes alveolar bones with tooth in it and the surrounding tissues, etc. Therefore, the Panoramic images we get have huge amount of data and high precision and. There are also much redundancy in these images. The area of Tooth from the whole image is about 50%, and the other areas are meaningless to the recognition system using tooth.

High-resolution makes the size of X-ray image is generally about 3000 * 1500. We need a very large space to store images (the average size of each image is approximately 2M bytes). And this can make it difficult to establish and maintain the large-scale database. It would also require a larger computer memory and longer time. In order to improve the effectiveness and efficiency of processing, we need to locate tooth position from original X-ray images.

Tooth positioning refers to locate the area of tooth by a rectangular box from the original X-ray image. There is not a separate subject for the research of the technology of tooth positioning at now. Some proposed teeth recognition systems use the technology complete artificial mark or artificial assisted semi-automatic mark to locate the tooth areas. Nomir put forward an algorithm of artificial assisted semi-automatic for tooth positioning. This method first mark the basic tooth areas manually, and then use a probability model to calculate the overall position of the tooth. And it achieves the position of each tooth precisely [3].

2 Algorithm of Tooth Positioning Based on Wavelet Transform and Edge Detection

In this paper, we proposed the algorithm of tooth positioning based on wavelet transform and edge detection. We get the feature patterns of both horizontal and vertical direction by analyzing dental X-ray images. And then we test the feature patterns respectively in both directions by using wavelet transform and edge detection so that we get the line which can separate upper and lower jaws and the rectangle which locate the area of tooth.

2.1 Definition of Wavelet Transform

Continuous wavelet transform is proposed by the Morlet and Grossman [11], for the function $\psi(x)$, if its fourier transform $\hat{\psi}(\omega)$ satisfies the equation (1), then $\psi(x)$ is the wavelet function.

$$\int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega = \int_{-\infty}^0 \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega = C_\psi < +\infty \tag{1}$$

In equation (1), $\psi(t)$ need to meet the requirements of the equation (2) in the domain of time.

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0 \tag{2}$$

Take the continuous wavelet transform for the function $f(t) \in L^2(\mathbb{R})$ as shown in equation (3)

$$(W_{\psi}f)(a, b) = |a|^{-1/2} \int_{-\infty}^{+\infty} f(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt \tag{3}$$

Stretching the value of the parameter S of wavelet function $\psi(t)$ transform to get the function $\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right)$. Equation 3 can also be expressed as

$$(W_{\psi}f)(a, b) = \int_{-\infty}^{+\infty} f(t) \overline{\psi_{a,b}(t)} dt = \langle f, \psi_{a,b} \rangle \tag{4}$$

If the function $f(x)$ satisfies the admissibility conditions of equation[1], then there exists its inverse transformation. And we can restore the original signal $f(t)$ accurately based on the wavelet transform $W_{\psi}f(a, b)$, the inverse transformation formula is expressed as equation 5.

$$f(t) = \frac{1}{C_{\psi}} \int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{a^2} W_{\psi}f(a, b) da db \tag{5}$$

and

$$C_{\psi} = \int_0^{+\infty} \frac{|\widehat{\psi}(\omega)|}{\omega} d\omega < +\infty \tag{6}$$

Daubechies wavelet is proposed by the French scholar Daubechies, , the function $\psi(t)$ is called the p level of Daubechies wavelet in the condition of Equation 7, and it can also be referred to db wavelets.

$$\int t^p \psi(t) dt = 0, p = 0, 1, 2, \dots, N \tag{7}$$

(1) finite support, which is affected by parameter N, The greater the value of N, the length longer.

(2) $\widehat{\psi}(\omega)$ is with N-order zero at $\omega = 0$ in the frequency domain.

(3) $\psi(t)$ and its displacement orthonormal integer that satisfies the equation [8] as shown below.

$$\int \psi(t) \psi(t - k) dt = \delta_k \tag{8}$$

The nature of Db wavelet makes it has the characteristics like smaller amount of calculation and flexible selection, so it has a wide range of applications in signal analysis.

2.2 Determine the Upper and Lower Boundaries of Area of Tooth by Using Wavelet Transform

In the X-ray images, as shown in figure 1(a) , the gray value of tooth area are generally high, which makes it separate from the background. And this characteristic can provide important basis of isolating tooth area accurately.

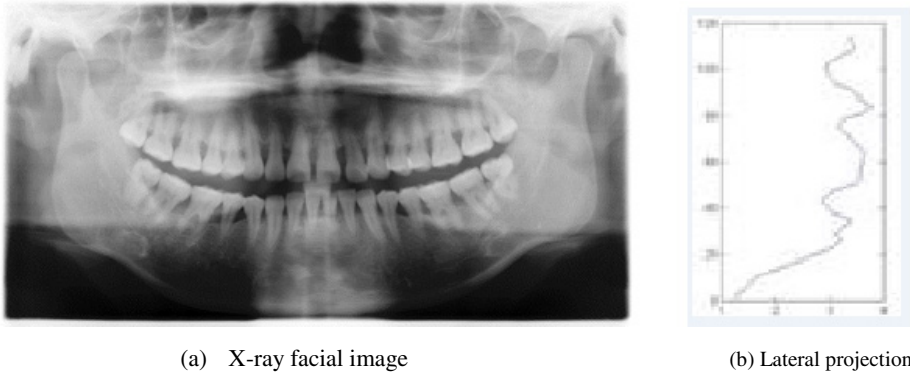


Fig. 1. X-ray image and its horizontal projection of the face

Another essential feature of the X-ray image is that tooth are all basically in the horizontal direction, so if we do the projection in horizontal direction, the value will be higher on the coordinates of corresponding to the tooth part due to the feature of high degree of tooth area. And the value of other parts are relatively lower, we can call it highland effect, as shown in figure 1(b).This provides possibilities for testing the tooth region.

From figure 1 we can find that there are three higher grayscale areas in the horizontal direction of the image. And the three regions can be corresponding to nostrils, maxillary tooth area and mandibular tooth area. In order to make sure that this characteristic has typical significance for all of the images, we make an analysis for 50 facial X-ray images. And we get the horizontal distribution trend diagram, as shown in figure 2 (b) (to facilitate observation, the figure is turned by 90 degree in horizontal direction).

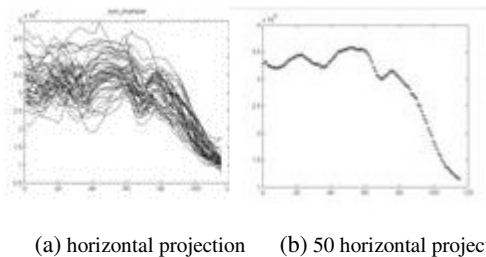


Fig. 2. X-ray images of facial horizontal projection and distribution trends

Figure2 (b) confirms our hypothesis that in the horizontal projection curve of facial X-ray image, the areas of nostrils, maxillary tooth and mandibular tooth are corresponding to the three areas with high value of the horizontal projection. There is a local minimum value between the maxillary tooth and mandibular tooth in the curve. And we can find that this local minimum value is corresponding to a black separation zone by analyzing the facial X-ray images.

As an effective time-frequency analysis method, wavelet transform can discover both the low and high frequency information through signal decomposition so that we can find something that we are interested in.

The projection will present a overall high-low trend as the horizontal projection of the facial X-ray image has a high-low effect. Therefore, we can discover the high-low distribution trend in the low frequency stage through the wavelet decomposition of the horizontal projection data and then find the corresponding position so as to locate our interested region—tooth area.

In the low frequency stage, the trend of the curve can be actually discovered through the wavelet transform of the horizontal projection, as shown in Figure 3. In this experiment, db10 wavelet [13] is adopted.

In the Level 4 approximation of the wavelet transform decomposition, the three regions are respectively corresponding to the three local maximums. And the basic regions of the maxillary tooth and mandibular can be affirmed by this way (which are respectively corresponding to the second and third local maximums).

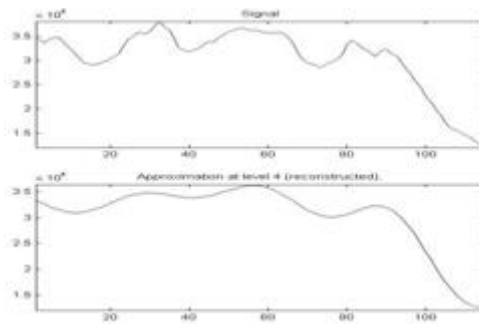
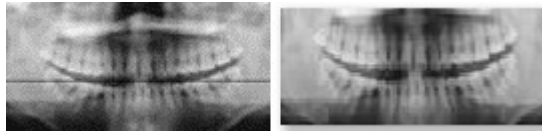


Fig. 3. Facial X-ray Image Horizontal Projection and Its WT Approximation at Level 4

As shown in Figure 1 (a) (b), the dividing line between the maxillary and the mandibular tooth is the local minimum between their corresponding horizontal projection regions. Hence, after the maxillary and mandibular teeth regions are found through the WT approximation at Level 4, their dividing line can be also found by finding the position of local minimum. The experiment result is shown as the solid black line in Figure 4 (a)



(a) Dividing line between maxillary tooth and mandibular tooth (b) Tooth area after upper and lower boundaries are affirmed.

Fig. 4. Dividing Line between Maxillary and Mandibular Teeth and Upper and Lower Boundaries

After the horizontal dividing line between the maxillary tooth and mandibular tooth is defined, the upper and lower boundaries of the tooth area can be also defined by one or two constants. These two empirical constants are respectively the deviation distance of the maxillary tooth and mandibular tooth from the middle dividing line. Many experimental tests have proved that a better result can be obtained when the upper and lower deviation distance are respectively set as 450 and 350, as shown in Figure 4 (b).

2.3 To Define Left and Right Boundaries of Tooth Position through Edge Detection

Edge detection is usually the first stage of the image processing and also one of the classical research topics in the machine vision field. Edge detection can significantly reduce the data volume of the image and reserve the structural information at the same time, so it can ease the calculation burden of the system substantially. The correctness and reliability of the detection result will have a direct influence on the understanding of the objective world from the perspective of machine vision system.

Traditional edge detection algorithm is mainly based on the gradient which is essentially the first order difference of the image. The pixels in the marginal area of the image changes more sharply so they have bigger gradients. Thus the gradient-based edge detection algorithm can achieve better effects when the interference of noise is small. Traditional edge detection is easy to understand and its calculated amount is small. As a result, it has been widely applied and developed in the early stage of image processing.

The calculation of the gradient of an image $f(x, y)$ can be shown in Equation 9.

$$\nabla f(x, y) = [G_x, G_y]^T = \left[\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y} \right]^T \tag{9}$$

The values of amplitude and phase can be obtained respectively after the gradient function of the image is obtained through the above calculation, as shown in Equations 10 and 11.

$$|\nabla f| = \left[\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2 \right]^{1/2} \tag{10}$$

$$\phi(x, y) = \arctan \left[\frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}} \right] \tag{11}$$

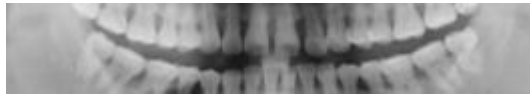
With these two equations, the amplitude and phase value of each pixel gradient can be respectively calculated by the convolution of the template and image. Then the two templates of G_x and G_y (respectively on behalf of the horizontal and vertical direction) can be combined to constitute a gradient-based edge detection operator. For such algorithms, there are differences among the scales and parameters of different templates, which could result in different functions. Among all the templates, Prewitt is the most representative one. As shown in Figure 5, Prewitt operator is an ideal edge template which can better detect the edge of an image.

1	0	-1
1	0	-1
1	0	-1

-1	-1	-1
0	0	0
1	1	1

Fig. 5. Prewitt Operator Template

The edge detection is often used to define the left and right boundaries of the teeth position. In this article, Prewitt edge detection algorithm is adopted.



(a) Narrow Zone between Maxillary and Mandibular Teeth



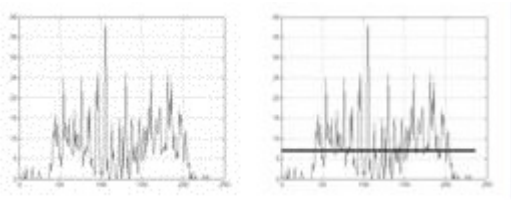
(b) Edge Detection Result

Fig. 6. Edge Detection Result of Teeth Zone

The dividing line between the maxillary and mandibular teeth has been obtained after the horizontal processing. Then center on the line and extend upward and downward for respectively 250 and 150 lines, thus a narrow horizontal zone is obtained. The edge detection result of this zone can be seen in Figure 6 ((a) is the narrow horizontal zone and (b) is its edge detection result).

As shown in Figure 6 (b), after the edge detection, there is more edge information in the teeth zone. However, there is almost zero information in non-teeth zone.

Through the vertical projection of the image after edge detection with the sum of each list of pixels as the projection value, the projection curve of the teeth zone can be obtained as shown in Figure 7 (a).



(a) Vertical Projection of Edge Image (b) Vertical Projection of Edge Image and Zone Average Value

Fig. 7. Edge Image Vertical Projection and Average Value

As shown in the vertical projection curve of the edge image, the projection value of the teeth position is higher and that of the two sides of teeth is lower. Except for few points, the projection value of the two sides is nearly zero. Therefore, the teeth zone in the image can be obtained through threshold processing and then the boundaries can be defined.

The author has designed a threshold calculation process, through which a valid threshold value can be obtained. The calculation steps are as follow:

- (1) Extract the maximum values in the edge vertical projection curve.
- (2) Rank the maximums and define the zone of the top 10 maximums.
- (3) Calculate the average of all non-maximums in the zone defined in Step (2).

Then this average is the threshold value.

The average value, i.e. the threshold value of the teeth position's vertical projection can be obtained through the above steps. This average can reflect the basic feature of the teeth position and effectively distinguish the teeth position and non-teeth position. The experiment result in Figure 7 (b) shows that, both the left side of the first point which is greater than the threshold value and the right side of the last point which is greater than the threshold value are non-teeth zone; the zone between these two points is exactly the teeth zone.

3 Experiment Result and Analysis

In order to verify the validity of the algorithm, the author has conducted experiments on 50 X-ray images, among which 41 images can output accurate results. Thus the accuracy rate of this algorithm reaches 82%. The experiment results demonstrate the effectiveness of the proposed algorithm. The experiment results of part of the images can be seen in Figure 8.

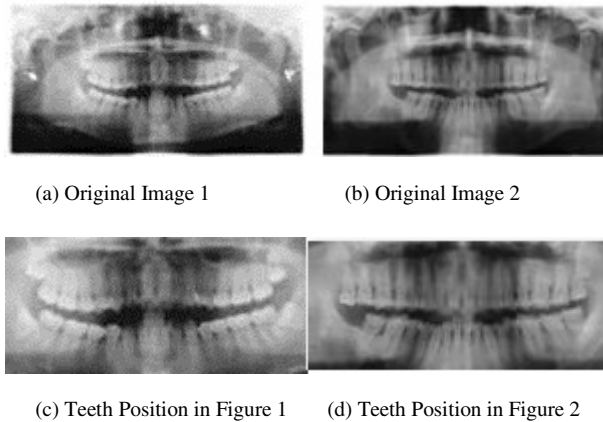


Fig. 8. Experiment Result

4 Conclusion

This article has put forward the teeth position detection algorithm for facial X-ray images based on wavelet transform and edge detection. In this algorithm, the facial X-ray image is projected horizontally; then the upper and lower boundaries of the teeth position is defined by analyzing the projection data with wavelet transform. After that, a narrow zone near the boundary is detected by edge detection method; the obtained edge image is projected vertically; then a threshold algorithm is designed to obtain the left and right boundaries of the teeth position. Through the horizontal and vertical processing, the teeth position in the facial X-ray image is defined. Compared with other algorithms, this algorithm to extract the teeth parts is more accurate, at the same time of reducing redundant information, it make the matching more effectively, and the method is simple, easy to implement. Plenty of simulation experiments have verified the accuracy and validity of this algorithm.

References

1. Yan, L., Zhang, H.: Individual identification of teeth and its application in forensic science. *Journal of Chinese People's Public Security University: Natural Science Edition* 12(003), 28–32 (2006)
2. Brannon, R.B., Morlang, W.M., Smith, B.C.: The gander disaster: dental identification in a military tragedy. *Journal of Forensic Sciences* 48(6), 1331–1335 (2003)
3. Jain, A.K., Chen, H.: Matching of dental X-ray images for human identification. *Pattern Recognition* 37(7), 1519–1532 (2004)
4. Nassar, D.E., et al.: Automatic construction of dental charts for postmortem identification. *IEEE Transactions on Information Forensics and Security* 3(2), 234–246 (2008)
5. Nomir, O., Abdel-Mottaleb, M.: Hierarchical contour matching for dental X-ray radiographs. *Pattern Recognition* 41(1), 130–138 (2008)

6. Nomir, O., Abdel-Mottaleb, M.: A system for human identification from X-ray dental radiographs. *Pattern Recognition* 38(8), 1295–1305 (2005)
7. Tan, Y., Tian, X., et al.: Research progress of identification using dental Imaging. *Criminal Technology* (001), 13–15 (2010)
8. Gao, D., Wang, Q., Ye, J., et al.: Full mouth dental surface layer of forensic identification of indicators. *Journal of Forensic Medicine* 24(002), 114–117 (2008)
9. Xu, B., Feng, H., Zhao, W., et al.: According to digital panoramic radio-graph tomographs individual identification studies. *Journal of Forensic Medicine* 25(005), 327–330 (2010)
10. Gao, D., Wang, H., Hu, J., et al.: Maxillofacial digital x-ray forensic identification. *Journal of Forensic Medicine* 22(1), 32–35 (2006)
11. Grossmann, A., Morlet, J.: Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis* 15(4), 723–736 (1984)
12. Yang, F.: *Engineering analysis and application of wavelet transform*. Science press (1999)
13. Daubechies, I.: Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics* 41(7), 909–996 (1988)

The Orientation Angle Detection of Multiple Insulators in Aerial Image^{*}

Zhenbing Zhao, Ning Liu, Mingxiao Xi, and Yajing Yan

School of Electrical and Electronic Engineering, North China Electric Power University,
Baoding, 071003, China

{zhaozhenbing2002, 15075230864,
yanyajing1013}@163.com,
578617034@qq.com

Abstract. Insulator is one of the important equipments on the transmission line, and its orientation angle detection is an important preprocessing step for accurate localization of insulator. This paper proposes a method of orientation angle detection of insulators in order to realize the orientation determination of multiple insulators in aerial image with complicated background. Firstly, extract sequential edge points and define their orientation angles of each linking contour in the preprocessed image. And secondly define the points with sign-changing angle as candidate points. Finally, the accurate points can be picked up by RANSAC (RANdom SAmple Consensus) from candidate points, and the straight line which accurate points locate in is the main direction of the target. Experimental results verify that the proposed method has higher detection accuracy compared with the existing methods that can lay the foundations for the localization of multiple insulators in complicated background.

Keywords: Orientation angle detection, multiple insulators, aerial image, complicated background, RANSAC.

1 Introduction

Insulator is indispensable equipment in the power system with the dual function of electrical insulation and mechanical support [1,2]. And the localization of insulator for monitoring its status is very necessary to avoid causing a failure of flashover and breakdown. The problem of insulators' orientation angle detection is finding the correct orientation so that the insulator would be in an upright position which makes insulators' localization easier and more accurate in aerial image with complex background.

There are a lot of methods of target orientation angle detection.

Literature [3,4] proposes a method that estimates the target aspect angular making use of track tracking. This method can realize the orientation angle estimation of moving target with high precision, but for still image, it may be helplessness.

^{*} The research is supported by "The Fundamental Research Funds for the Central Universities" under grant number 12MS122.

Literature [5] develops an effective algorithm for human face orientation detection applying the symmetry features of the human face. It is effective and promising for human face orientation detection, but, it may be not fit for generic images without symmetry. Literature [6,7] can solve the identification of the orientation of sample polygons in bizarre circumstance effectively. This method has a good stability and generality, and high robust for the identification of sample polygon in all cases, but the accuracy would be reduced significantly for image with curve edges and complicated background. Literature [8] proposes a recognition method to deal with multi-view targets in infrared images synthetically utilizing image invariant moments feature and SVM (Support Vector Machine) classification method. But this method can solve the orientation detection of targets without background, and the division of view-Angle coverage increases the error in the determination of orientation angle, and also the complexity of SVM classification method would bring large calculation. In literature [9], a novel skew correction algorithm is proposed focusing on boundary line that optimizes speed and accuracy by Hough transform to get the skew corrected license plate image. But this method achieves high accuracy of orientation angle detection at the expense of large calculation, and for images with complex background, the calculation would be increasing when the target area is small which causes bad angle detection. Literature [10] develops a probabilistic approach to image orientation detection by confidence-Based integration of low-Level and semantic cues within a Bayesian framework. The computation is increasing caused by Bayesian, especially in complex background. Literature [11] presents a document skew and orientation detection technique by white run histograms through scanning documents in horizontal and vertical directions. This method is capable of detecting arbitrary skew orientation of documents with high speed and low computation. Literature [12] introduces a generic, scale-Independent algorithm which is capable of accurately detecting the global skew angle of document images within the range $[-180^\circ, 180^\circ]$. Despite its generality, the method is very fast and requires no explicit parameters with high accuracy and robustness. But literature [11,12] would achieve high accuracy because of a large amount of equidistant interline spacings leading to significant limitations, and would fail on complex images containing almost exclusively majuscules, digits and mathematical formulae. Literature [13] presents a recognition driven method for page orientation detection. A small subset of text lines are extracted in the four orientations, and then the outputs of the OCR (Optical Character Recognition) are evaluated to deduce the right orientation. The method can handle documents containing small lines of text, and it is able to detect multiple orientations. But it may not realize the detection of arbitrary angle. Literature [14] proposes a script identification method that works for unknown orientation for all official Indian scripts. It has high accuracy, but classifier may result in large computation. Literature [15] uses PCA (Principal Component Analysis) to realize insulators' tilt correction, and it has a high accuracy for insulator images without background, but it may be not effective for images with complex background.

The above methods can realize the orientation angle detection, but they all have significant limitations, and can not realize the orientation determination of multiple insulators in aerial image with complicated background. This paper proposes a method of orientation angle detection of insulator in order to realize the orientation angle detection of multiple insulators in aerial image with complicated background. Firstly, extract sequential edge points and define their orientation angles of each linking contour in the preprocessed image. And secondly define the points with sign-Changing angle as candidate points. Finally, the accurate points can be picked up by RANSAC from candidate points, and the straight line which accurate points locate in is the main direction of target. Experimental results verify that the proposed method has higher detection accuracy compared with the existing methods by using a large number of insulator images that can lay the foundations for the localization and fault diagnosis of insulators in complex background.

2 Proposed Method

This paper proposes a method of orientation angle detection of insulator in order to realize the orientation angle detection of multiple insulators in aerial image with complicated background. The flow chart of orientation angle detection of insulators is shown in Fig. 1.

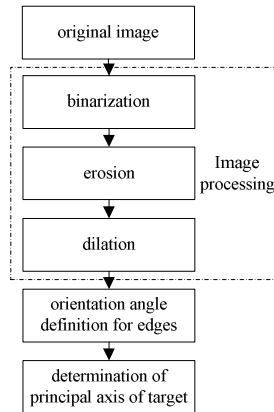


Fig. 1. The flow chart of orientation angle detection of insulator

2.1 Image Preprocessing

2.1.1 Image Binarization

Convert the original aerial image into binary image by threshold segmentation to realize the separation of foreground and background. Take aerial insulator image (Im 1) as an example, the foreground can be gotten which is shown in Fig. 2.

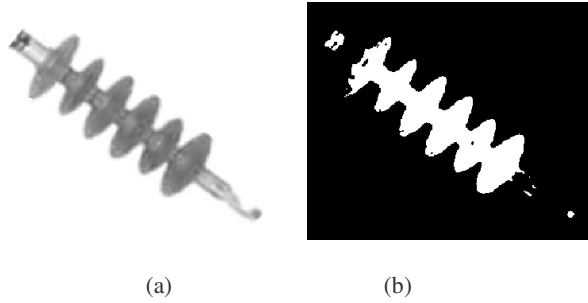


Fig. 2. Im 1 and its binary image. (a) Im 1. (b) Its binary image

2.1.2 Morphological Filtering

Shown in Fig. 2(b), there are a lot of internal holes inside the insulator, and the edges are blurry which lead to difficult edge extraction. This paper adopts double cascade filtering method to erode and dilate the binary image. And after morphological filtering, internal holes of Im 1 are filled basically, and the edges become smoothness and clearness in Fig. 3(a).

There are also some residual small regions after filtering seen from Fig. 3(a). Set a threshold and remove the regions whose areas are less than the value shown as Fig. 3(b) that improves accuracy of the detection of insulator's orientation angle and simplifies the computation.

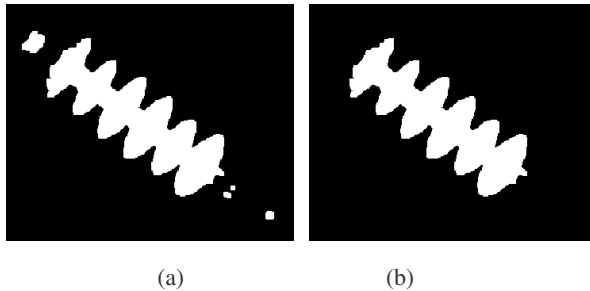


Fig. 3. Results of preprocessing of Im 1. (a) Result of morphological filtering of Im 1. (b) Result of preprocessing of Im 1.

2.2 Orientation Angle Definition for Edges

Extract the contours of Im 1 shown in Fig. 4(a), and link edge pixels together into lists of sequential edge points, one list for each edge contour. A contour starts/stops at an ending or a junction with another contour. And the contour would be discarded if it is less than a specific value.

The contour of binary image is concave-Convex and irregular curve after preprocessing, which brings a lot of trouble for the orientation angle detection of insulator. This paper uses a large number of straight line segments with specified tolerance to

approximate the irregular curve contour shown in Fig. 4(b). Thus each contour is composed by plenty straight line segments and the points in each contour can be formed as a list of connected edge points.

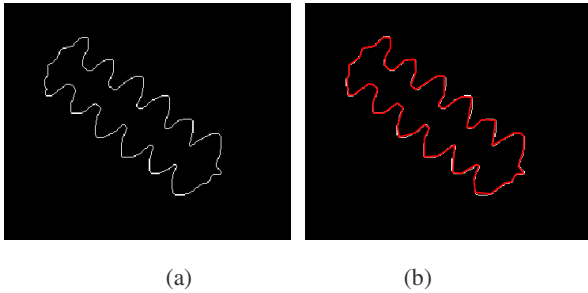


Fig. 4. Curve edge of Im 1 and processed edge. (a) Curve edge of Im 1. (b) Processed edge.

Define the orientation angle for each point in each list, and all the angles are composed of a set of sequential angular variation. The points with sign-Changing angle would be extracted as candidate points. If the number of candidate points of the contour is too less, the contour can be ignored as pseudo-insulator.

Shown in Fig. 5, *a*, *b*, *c* is three adjacent edge points in the list of the contour. Define the angle of vertical direction is 0° , then the angle of *a* is -45° and the angle of *b* and *c* is 45° , thus *b* is the points with sign-Changing angle which can be defined as candidate point. And Fig. 6 shows candidate points of each connected contour in Im 1.

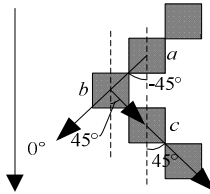


Fig. 5. The schematic of orientation angles of edge points

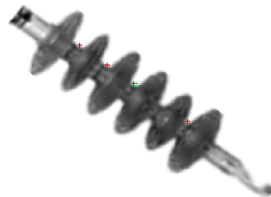


Fig. 6. Candidate points of Im 1

2.3 The Determination of Main Direction of Insulator

All the candidate points can be extracted including some error points, and the accurate points can be picked up by RANSAC, and the straight line which points locate in is the main direction orientation of insulator shown in Fig. 7.

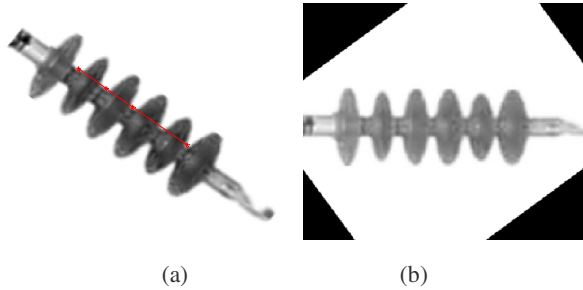


Fig. 7. The detected orientation angles of insulator. (a) Main direction of insulator. (b) Rotated insulator.

3 Experimental Results

3.1 The Results of Angles Detection of Multiple Insulators

The results of angles detection of multiple insulators are as following:

Fig. 8(a) is an image with two insulators (Im 2), and (b) is its binary image after preprocessing. It can be seen that the edges of binary image become smoothness and clearness which simplifies the computation and increases the accuracy.

Extract the contours of Im 2 shown in Fig. 9(a), and link edge pixels together into lists of sequential edge points for every contour of Im 2. Reserve the contours which are longer than definite value.

For each contour, the steps of the detection of orientation angle are as following:

Shown in Fig. 9(b), the contour of binary image which is concave-Convex and irregular curve after preprocessing is approximated by a large number of line segments with specified tolerance, and each color represents a connected contour composed by straight line segments. Thus the points in each contour can be formed as a list of connected edge points.

Compute the orientation angle for each point in each list, and all the angles are composed of a set of sequential angular variation. The points with sign-Changing angle would be extracted as candidate points. If the number of candidate points of the contour is too less, the contour can be ignored as pseudo-insulator. Fig. 10 shows candidate points of each connected contour in Im 2.

All the candidate points can be extracted including some error points, and the accurate points can be picked up by RANSAC, and the straight line which points locate in is the main direction of insulator.

After iterations, the orientation angles of all the connected contours can be detected finally. Fig. 11 is the results of orientation angle detection of all insulators.

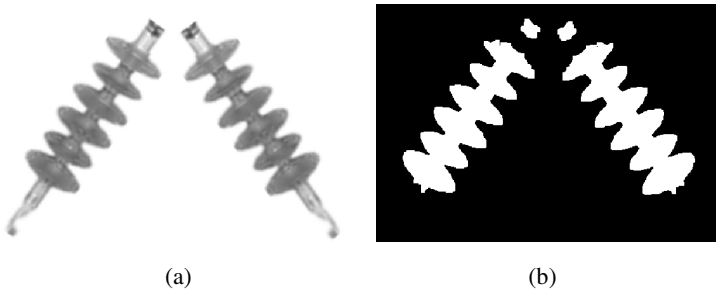


Fig. 8. Im 2 and its binary image after preprocessing. (a) Im 2. (b) Its binary image after preprocessing.

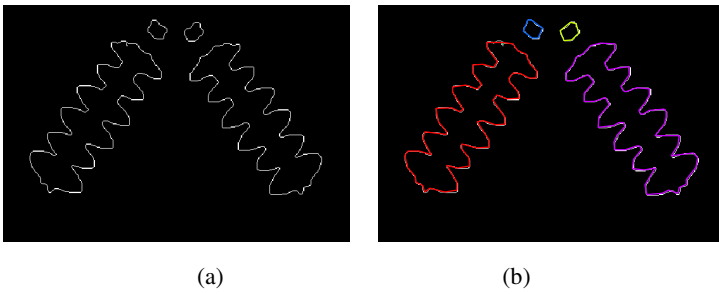


Fig. 9. Curve edge of Im 2 and processed edge. (a) Curve edge of Im 2. (b) Processed edge.



Fig. 10. Candidate points of each connected contour in Im 2. (a) Candidate points of contour 1. (b) Candidate points of contour 2.

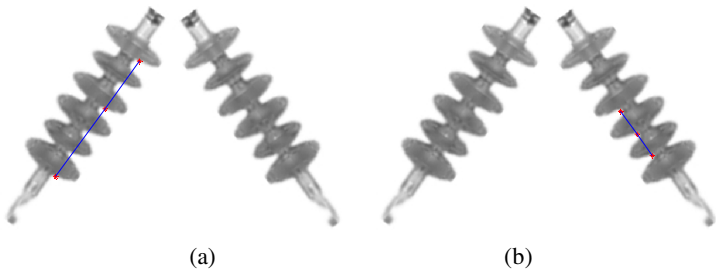


Fig. 11. The detected orientation angles of all insulators. (a) The detected orientation angles of insulator 1. (b) The detected orientation angles of insulator 2.

3.2 The Results of Angles Detection of Insulators in Complex Background

Take aerial insulator image with complex background (Im 3) as an example, the results of angle detection of insulator are as following:

In Fig. 12, (a) is Im 3 with complex aerial background including tower, power lines and vegetation, and (b) is its binary image with smooth and clear edges after preprocessing.

Extract the contours of Im 3 shown in Fig. 13(a). The concave-Convex and irregular curve contour would be approximated by a large number of line segments shown in Fig. 13(b). Thus the points in each contour can be formed as a list of connected edge points.

Compute the angle of each point and all the angles are composed of a set of sequential angular variation, and define the points with sign-Changing angle as candidate points. Fig. 14(a) shows candidate points of each connected contour in Im 3. All candidate points can be extracted including some error points, and the accurate points can be picked up by RANSAC shown in Fig. 14(b), and the straight line which points locate in is the main direction of insulator.

After iterations, the orientation angles of all the connected contours can be detected finally.

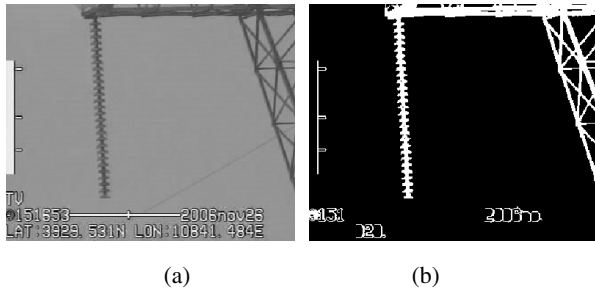


Fig. 12. Im 3 and its binary image after preprocessing. (a) Im 3. (b) Its binary image after preprocessing.

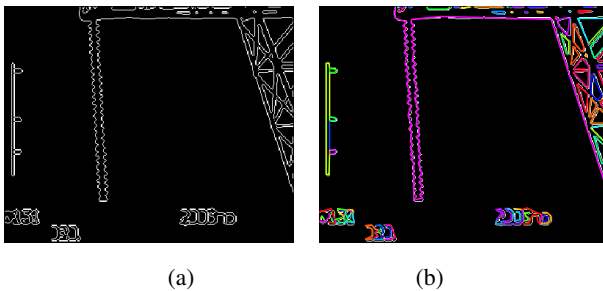


Fig. 13. Curve edge of Im 3 and processed edge. (a) Curve edge of Im 3. (b) Processed edge.

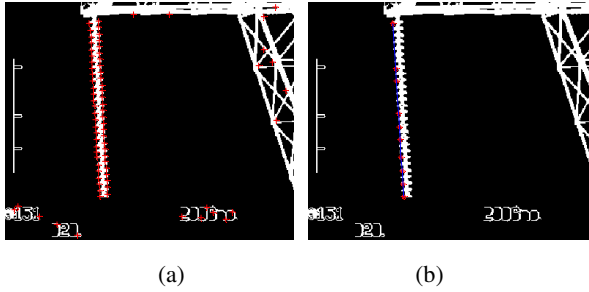


Fig. 14. Candidate points of each connected region in Im 3. (a) Candidate points of region 1. (b) The detected orientation angles of all insulators.

3.3 The Comparison of Existing Orientation Angle Detection Methods and Proposed Method

In method 1 [16], rotate the target image to the largest of its horizontal projection, and then the rotational angle is defined as the orientation angle of target. But the obtained optimum tilt angle is an optimization procedure which needs multiple projections to approximate to that leads to a large calculation. In method 2, the orientation of target would be considered as the main direction when it has the maximum number of pixels, but it is easily susceptible by noises. In method 3, radon transform [17] in the range of $[-20^\circ, 20^\circ]$ is implemented on the target image. Then the accumulative total of absolute value of difference of the results is calculated, respectively. The angle of the radon transform corresponding to the maximal accumulative total is confirmed as a skew angle of target image. But it may be not able to obtain the accurate tilt angle. Method 2 affirms the orientation angle of target by Hough [18], but there is a contradiction between accuracy and computation complexity. The above methods are dealing with target without background, and they may not meet the demand of orientation angle detection of complex image.

This paper firstly rotates the original image from 0° to 60° at 10° intervals, and estimates its rotation angle by the above four methods and proposed method to verify the effectiveness of our method.

From the detection angle of method 1 in Table 1, the performance of method 1 is reducing while the rotation angle is increasing. The horizontal projection is equal when the two rotation angles of original image are symmetrical about horizontal direction, and this method can not make correct judgment. Method 2 contains notable error caused by noises. And compared with method 3 and 4, the detected angles of the proposed method are closer to the correct value; the method 3 and 4 are susceptible to noises and prone to be lost in local maxima, and the proposed method is robust by noises which can produce high accuracy. In conclusion, the proposed method is more approximate to the exact orientation angle and has higher accuracy within a limit of allowable error.

Table 1. The comparison of existing methods and proposed method

Rotation angle	Method 1	Method 2	Method 3	Method 4	Proposed method
0.00	3.00	42.70	1.00	1.00	0.00
10.00	9.00	16.44	11.00	9.00	9.17
20.00	19.00	26.50	17.00	19.00	19.32
30.00	29.00	28.76	33.00	29.00	29.00
40.00	40.00	38.60	36.00	39.00	41.74
50.00	39.00	48.46	48.00	49.00	49.43
60.00	29.00	53.04	58.00	59.00	60.11

4 Conclusions

This paper proposes a novel method of orientation angle detection of aerial multiple insulators in complex background. It extracts sequential edge points and defines their orientation angles for each linking contour in the input image, and defines the points with sign-changing angle as candidate points, and the accurate points can be picked up by RANSAC from candidate points, the straight line which accurate points locate in is the main direction orientation of insulator. Experimental results verify that the proposed method has higher detection accuracy compared with the existing methods that can lay the foundations for the localization of multiple insulators in complex background.

Reference

1. Zhao, J.J., Liu, X.T., Sun, J.X., Lei, L.: Detecting Insulators in the Image of Overhead Transmission Lines. In: Huang, D.-S., Jiang, C., Bevilacqua, V., Figueroa, J.C. (eds.) ICIC 2012. LNCS, vol. 7389, pp. 442–450. Springer, Heidelberg (2012)
2. Yan, S.J., Jin, L.J., Zhang, Z., Zhang, W.H.: Research on Fault Diagnosis of Transmission Line Based on SIFT Feature. In: Guo, C., Hou, Z.-G., Zeng, Z. (eds.) ISNN 2013, Part II. LNCS, vol. 7952, pp. 569–577. Springer, Heidelberg (2013)
3. Liu, X.M., Shen, J.F.: Research on Getting Aspect Angular for Radar/Infrared-Imaging ASM. *Journal of Naval Aeronautical and Astronautical University* 24(3), 304–306 (2009)
4. Liu, J.M., Wu, Z.L., Wang, J.: Method to Estimate the Trajectory of Space Target with Infrared Mono-station. *Laser & Infrared* 41(10), 1167–1171 (2011)
5. Han, S., Pan, G., Wu, Z.: Human Face Orientation Detection Using Power Spectrum Based Measurements. *IEEE Transactions on Automatic Face and Gesture Recognition*, 791–796 (2004)
6. Ding, J., Jiang, N., Rui, T.: Robust Algorithm for Identifying the Orientation of Simple Polygons. *Journal of Computer-aided Design & Computer Graphics* 17(3), 442–447 (2005)
7. Ding, J., Jiang, N., Rui, T.: A New Algorithm for Identifying Polygons' Orientation. *Computer Engineering* 32(9), 47–50 (2005)

8. Ma, C.J., Li, X.X., Yang, H., Wu, D., Wang, J.W.: Multi-view Target Recognition Algorithm Based on Support Vector Machine Classification. *Laser & Infrared* 39(1), 88–91 (2009)
9. Arulmozhi, K., Perumal, S.A., Priyadarsini, C.S.T., Nallaperumal, K.: Image Refinement Using Skew Angle Detection and Correction for Indian License Plates. In: 2012 IEEE International Conference on Computational Intelligence & Computing Research, pp. 1–4. IEEE Press, Coimbatore (2012)
10. Luo, J., Boutell, M.: Automatic Image Orientation Detection Via Confidence-based Integration of Low-level and Semantic Cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(5), 715–726 (2005)
11. Lu, S.J., Wang, J., Tan, C.L.: Fast and Accurate Detection of Document Skew and Orientation. In: 9th International Conference on Document Analysis and Recognition, Parana, vol. 7, pp. 684–688 (2007)
12. Konya, I., Eickeler, S., Seibert, C.: Fast Seamless Skew and Orientation Detection in Document Images. In: 20th International Conference on Pattern Recognition, Istanbul, pp. 1924–1928 (2010)
13. Rangoni, Y., Shafait, F., Van Beusekom, J., Breuel, T.M.: Recognition Driven Page Orientation Detection. In: 16th IEEE International Conference on Image Processing, pp. 1989–1992. IEEE Press, Cairo (2009)
14. Ghosh, S., Chaudhuri, B.B.: Composite Script Identification and Orientation Detection for Indian Text Images. In: 2011 International Conference on Document Analysis and Recognition, Beijing, pp. 294–298 (2011)
15. Li, B.F., Wu, D.L., Cong, Y., Xia, Y., Tang, Y.D.: A Method of Insulator Detection from Video Sequence. In: 2012 International Symposium on Information Science and Engineering, Shanghai, pp. 386–389 (2012)
16. Wang, M., Wang, G.H.: An Optimization Algorithm of Vehicle License Plate Correction Based on Minimum Projection Distance. In: 9th International Conference for Young Computer Scientists, Hunan, pp. 1701–1705 (2008)
17. Jia, X.D., Li, W.J., Wang, H.J.: Novel Approach for Vehicle License Plate Tilt Correction Based on Radon Transform. *Computer Engineering and Applications* 44(3), 245–248 (2008)
18. Wang, L.H., Wang, J.L., Liang, Y.H.: Hough Transform and Its Application in Declining License Plate. *Information and Electronic Engineering* 2(1), 45–48 (2004)

Improved Robust Watermarking Based on Rational Dither Modulation

Zairan Wang^{1,2}, Jing Dong¹, Wei Wang¹, and Tieniu Tan^{1,2}

¹ Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences

² College of Engineering and Information Technology,
University of Chinese Academy of Sciences
{zairan.wang, jdong, wwang, tnt}@nlpr.ia.ac.cn

Abstract. Rational dither modulation (RDM) watermarking was presented to resist amplitude scaling attack. This property is achieved by quantizing the ratio of consecutive samples instead of samples themselves. In this paper, we improve the performance of basic RDM watermarking to resist more types of watermarking attacks. We improve the robustness of our modified RDM watermarking by the following three aspects: 1) The quantization step size is increased by modifying two coefficients instead of only one coefficient in the basic RDM method, 2) Several modification rules are defined to reduce embedding distortion, and 3) The coefficients with larger magnitudes in the lowest sub-band in DWT domain are selected to embed watermark. A variety of attacks are implemented to evaluate the performance of our method. Experimental results demonstrate that our method outperforms the basic RDM method and two state-of-the-art watermarking methods over a wide range of attacks and it also has good imperceptibility.

Keywords: Watermarking, RDM, Amplitude scaling attack.

1 Introduction

Digital watermarking has always drawn extensive attention for digital copyright protection since it was born. So far, many watermarking schemes have been proposed in the literature. One of the most popular algorithms is quantization based watermarking scheme [1]. The main idea of quantization based watermarking is that the host data is quantized into different quantization intervals according to different watermark information. Chen and Wornell [1] proposed a quantization based watermarking scheme which they called quantization index modulation (QIM). Chen [2] quantized the mean of a set of wavelet coefficients to embed watermark. Lin [3] embedded watermark by quantizing the local maximum coefficients in mid-frequency wavelet sub-band. Chen and Horng [4] embedded watermark by modulating the wavelet coefficients.

The main weakness of QIM based watermarking is that it is very sensitive to amplitude scaling attack. Therefore, many watermarking schemes have been proposed to deal with this problem in recent years. Shterev [5] proposed a maximum likelihood technique to estimate the amplitude scale in the watermark extraction process. Some researchers made use of amplitude-scale invariant codes to combat amplitude scaling

attack [6,7]. Moreover, some amplitude-scale invariant features were used to embed watermark. In the angle QIM (AQIM) [8], the angle of a vector of image samples was quantized. Zhu introduced a normalized dither modulation (NDM) [9]. The main idea of NDM was to construct a gain-invariant vector with zero mean for quantization. Nezhadarya proposed the gradient direction watermarking (GDWM) [10], where the direction of gradient vectors was uniformly quantized.

Fernando proposed an alternative QIM method [11], called rational dither modulation (RDM), where a gain-invariant adaptive quantization step size at both embedder and decoder was used to against gain attacks. Inspired by their previous work, we propose an improved version of the basic RDM in DWT domain to obtain better robustness, because robustness is a basic requirement for watermarking used for copyright protection. To this aim, we improve the basic RMD watermarking algorithm mainly in the following three aspects: First, two coefficients instead of only one coefficient in the basic RDM are modified to embed watermark, then the allowed quantization step size can be increased. Second, several modification rules are defined to reduce embedding distortion and to improve robustness. In addition, significant coefficients in DWT domain are selected to embed watermark, because they are more robust to resist various kinds of attacks. A wide range of attacks are tested to evaluate the performance of our method, such as amplitude scaling, image filtering, JPEG compression, noise addition, rotation and resizing. We can see that our method is not only robust to amplitude scaling attack but also robust to common signal processing attacks. We compare our method with the basic RDM watermarking method [11] and two state-of-the-art watermarking methods proposed in [10] and [13]. Experimental results demonstrate that our method outperforms the three compared watermarking methods.

The rest of this paper is structured as follows. Section 2 introduces the details of RDM watermarking method. Section 3 presents the proposed watermarking method. Then, experimental results are shown in Section 4. Conclusions are given in Section 5.

2 Improved RDM Watermarking

2.1 Basic RDM Watermarking

RDM watermarking as a QIM approach was first proposed by Gonzalez [11] to against amplitude scaling attack. The quantization step size of RDM can be seen as a variable step quantizer, whose size is a function of several past watermarked samples. In this paper, samples denote as coefficients in the lowest sub-band in DWT domain. Then a gain invariant adaptive quantization step size is obtained at both embedder and decoder.

In the basic RDM, the set of rational functions $g : \mathbb{R}^L \rightarrow \mathbb{R}$, $L \geq 1$ are used, which have the property that:

$$g(\rho y) = \rho g(y), \text{ for all } \rho > 0, y \in \mathbb{R}^L. \quad (1)$$

Given a host signal vector, $x = (x_1 \dots x_N)$ and a watermarked signal vector, $y = (y_1 \dots y_M)$, then the k th bit $m_k \in \{0, 1\}$ of a watermark message is embedded in the L th-order RDM as:

$$y_k = g(y_{k-L}^{k-1}) Q_{m_k} \left(\frac{x_k}{g(y_{k-L}^{k-1})} \right) \quad (2)$$

where y_{k-L}^{k-1} denotes the set of watermarked samples $(y_{k-L} \dots y_{k-1})$ and L is the number of previous watermarked samples used to calculate the function $g()$, the function $Q_{m_k}(\cdot)$ is the standard quantization operation, so that the quantized samples belong to the shifted lattices:

$$Q_{m_k}(\cdot) = \begin{cases} 2\Delta\mathbb{Z} & \text{if } m_k = 0 \\ 2\Delta\mathbb{Z} + \Delta & \text{if } m_k = 1 \end{cases} \quad (3)$$

where Δ is the fixed quantization step size.

At the decoding side, suppose z_k is a possibly distorted sample. The hidden bit is recovered by applying standard quantization decoding procedure to the ratio between z_k and its previous samples z_{k-L}^{k-1} :

$$\hat{m}_k = \arg \min_{m_k} \left| \frac{z_k}{g(z_{k-L}^{k-1})} - Q_{m_k} \left(\frac{z_k}{g(z_{k-L}^{k-1})} \right) \right|, m_k \in \{0, 1\} \quad (4)$$

As to the choice of $g()$, a very large possible functions can be chosen, including the l_p -norms, given by:

$$g(y_{k-L}^{k-1}) = \left(\frac{1}{L} \sum_{i=1}^L |y_{k-i}|^p \right)^{1/p} \quad (5)$$

In this paper, the l_1 norm is considered, as in [11] and [12].

2.2 Improved RDM Watermarking

A weakness of the basic RDM algorithm is that attacking noise has big influence on the decoding quantization step size, though the influence can be decreased by increasing L . Hence, we modify the basic RDM algorithm to increase the quantization step size, then better robustness can be obtained. In the basic RDM algorithm, a ratio is computed using a un-watermarked sample and several past watermarked samples, thus only the un-watermarked sample can be modified. Different from the basic RDM method, we compute a ratio of two un-watermarked samples, thus two samples can be modified simultaneously to embed watermark, which increases the quantization step size.

Let $x_i \in R, i = 1, 2$ be two samples, the ratio of them r_x is computed as:

$$r_x = \frac{\min(x_1, x_2)}{\max(x_1, x_2)} \quad (6)$$

Obviously, r_x is in the range of 0 and 1. The watermarked ratio r_y is quantized as follow:

$$r_y = Q_{m_k}(r_x), m_k \in \{0, 1\} \quad (7)$$

where $Q_{m_k}(\cdot)$ is the quantization function, which is defined as:

$$Q_{m_k}(r_x) = \begin{cases} \Delta \lceil \frac{r_x}{\Delta} \rceil & \text{if } \text{mod}(\lceil \frac{r_x}{\Delta} \rceil, 2) = m_k \\ \Delta \lceil \frac{r_x}{\Delta} \rceil + \Delta & \text{if } \text{mod}(\lceil \frac{r_x}{\Delta} \rceil, 2) \neq m_k \text{ and } \Delta \lceil \frac{r_x}{\Delta} \rceil \geq r_x \text{ or } r_x = 0 \\ \Delta \lceil \frac{r_x}{\Delta} \rceil - \Delta & \text{if } \text{mod}(\lceil \frac{r_x}{\Delta} \rceil, 2) \neq m_k \text{ and } \Delta \lceil \frac{r_x}{\Delta} \rceil < r_x \text{ or } r_x = 1 \end{cases} \quad (8)$$

where $[\cdot]$ is the round function, and $\text{mod}(\cdot)$ denotes the modulo function. It is easy to see that the watermarked ratio r_y is an even or odd multiple of Δ .

To get the watermarked ratio r_y , we modify x_1 and x_2 to y_1 and y_2 respectively. Suppose x_2 is larger than x_1 , then the following equation must be satisfied:

$$r_y = \frac{y_1}{y_2} = \frac{x_1 + d_1}{x_2 + d_2} \tag{9}$$

where d_1 and d_2 are the modification strength of x_1 and x_2 , respectively.

In watermarking algorithms, robustness and transparency are always two conflicting factors. It is generally accepted that high transparency will decrease robustness and high robustness will limit transparency on the other hand. So there must be a tradeoff between the two factors. In our scheme, at a given quantization step size, we want that the embedding distortion which results from the sample modification will be as small as possible. To this aim, we define several modification rules as follows:

- Decrease x_2 and increase x_1 , if r_y is larger than r_x ;
 - Increase x_2 and decrease x_1 , if r_y is smaller than r_x ;
 - The amount of modification of x_2 should be larger than the modification of x_1 .
- Because it is widely accepted that larger coefficients allow greater modification strength.

To satisfy the above modification rules, we let d_1 and d_2 meet the following equation:

$$d_1 = -\frac{x_1}{x_2}d_2 \tag{10}$$

Combined with (9) and (10), d_1 and d_2 can be calculated as:

$$d_1 = -\frac{x_1^2 - x_1x_2r_y}{x_1 + x_2r_y}, \quad d_2 = \frac{x_1x_2 - x_2^2r_y}{x_1 + x_2r_y} \tag{11}$$

Afterwards watermarked samples y_i are obtained. At the decoding end, the watermarked signal y may be attacked and changed to z . The watermark bit \hat{m}_k is decoded by the minimal distance decoder:

$$\hat{m}_k = \arg \min_{m_k} ||r_z - Q_{m_k}(r_z)||, \quad m_k \in \{0, 1\}. \tag{12}$$

Now, let us see why our method can increase the quantization step size. As previously said, only one sample can be modified in the basic RDM algorithm, but two samples can be modified in our method. Without loss of generality, we suppose that x_1, x_2, d_1, d_2 meet the following conditions:

$$x_1 = x_2, \quad d_1 = -d_2 \tag{13}$$

As l_1 norm is used, we can suppose that the function of past watermarked samples $g(y_{k-L}^{k-1})$ is approximately equal to x_2 . It is clear that the following inequality is satisfied:

$$\left| \frac{x_1 + d_1}{x_2 + d_2} - \frac{x_1}{x_2} \right| > \left| \frac{x_1 + d_1}{x_2} - \frac{x_1}{x_2} \right| \approx \left| \frac{x_1 + d_1}{g(y_{k-L}^{k-1})} - \frac{x_1}{g(y_{k-L}^{k-1})} \right| \tag{14}$$

Thus the quantization step size in our method is larger than that of the basic RDM watermarking method.

3 Proposed Watermarking Methods

We implement our improved method in wavelet domain. The lowest frequency sub-band is selected, because it is the perceptually significant region which is robust enough to resist various attacks. The following are the details.

1) *Preprocessing*: The significant coefficients, which have large magnitude, are chosen to embed watermark, because they are more robust and the allowed modification strength of them is larger, which makes the embedding more robust to attacks. In our scheme, we divide the lowest frequency sub-band into non-overlapping blocks, and select the largest two coefficients from each block to embed one bit. However, in natural images, some parts do not contain any significant coefficient, which are improper for embedding, or some parts contain more than two significant coefficients, which may cause a waste. In other words, the significant coefficients are not uniformly distributed. To settle this problem, before partitioning the lowest sub-band into blocks, we first scramble the selected sub-band, so that the order of the coefficients will be disrupted and significant coefficients are distributed more uniformly.

2) *Watermark embedding*: The watermark embedding procedure is illustrated in Fig. 1, which can be described as following steps:

1. D level DWT is applied on the host image.
2. The lowest frequency sub-band is selected and scrambled using a secret key K .
3. Divide the scrambled sub-band into non-overlapping blocks with size of w .
4. Select the largest two coefficients from each block, then quantize the ratio of them to embed a watermark bit as introduced in Section 2.2.
5. Finally, the scrambled sub-band is descrambled and inverse discrete wavelet transform is applied, then the watermarked image is generated.

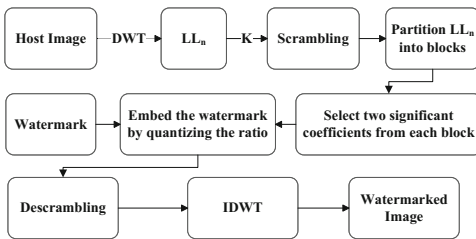


Fig. 1. Flowchart of watermark embedding

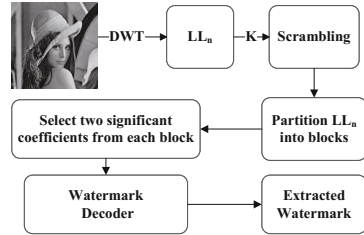


Fig. 2. Flowchart of watermark decoding

3) *Watermark decoding*: The process of watermark decoding is illustrated in Fig. 2, which can be described as follows:

1. The watermarked image is decomposed with D level discrete wavelet transform, then the lowest frequency sub-band is scrambled with the secret key K and divided into non-overlapping blocks with size of w , as described in the first three steps in the watermark embedding process.
2. Select the largest two coefficients from each block, denoted as z_1 and z_2 . Then decode the watermarked bit by Equation (12).

4 Experimental Results

In our experiments, various attacks are tested to evaluate the robustness of our method, including amplitude scaling, image filtering, adding noise, rotation and resizing. All the test images are of size 512×512 and in gray-scale. All test images are decomposed with three level wavelets, and the block size w is 4. The peak signal-to-noise ratio (PSNR) is used to measure the similarity of the original image to the watermarked image. And the bit error rate (BER) is used to judge the existence of the watermark.

4.1 Comparison with Basic RDM

In this section, we compare the improved RDM algorithm with the basic RDM method. We denote the basic RDM algorithm as RDM-Basic. The RDM-Basic is implemented with 10th, 30th and 50th order respectively. The watermark is a 1024 bits Gaussian pseudo-random sequence. To fill up the capacity, the watermark is embedded with four times repeatedly in RDM-Basic method, while in the proposed improved RDM method, the watermark is embedded only once. The test images are "Peppers," "Baboon," "Barbara," and "Lena." For fair comparison, the PSNR values of all the images are kept consistent (about 42dB) for the two watermarking algorithms. We repeat our method 100 times with 100 different pseudo-random binary watermarks. The BER is calculated by averaging the results of 100 times of the four test images. Fig. 3 shows two watermarked images of the proposed watermarking algorithm. From the figure, we can see that there is no visual distortion of the watermarked images.

Lossy JPEG compression is the most commonly image process in applications. The results of robustness comparison against this attack are shown in Fig.4. It is clearly demonstrated that the proposed improved RDM method outperforms RDM-Basic method under JPEG compression attack, especially when the quality factor belows 30.

Fig. 5 shows the BER comparison under amplitude scaling attack. It can be seen that both our proposed method and the RDM-Basic method are very robust to this attack as expected. The nonlinear amplitude scaling, gamma correction, is also tested. The results are shown in Table 1. It can be seen that the proposed improved RDM method can significantly improve the performance against gamma correction.



Fig. 3. Test images

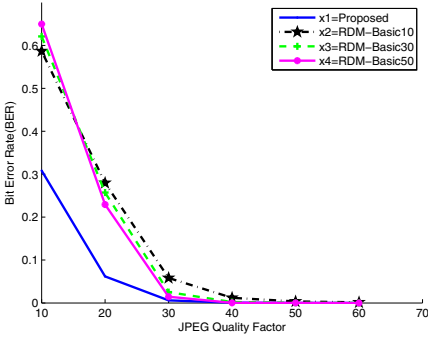


Fig. 4. BER under JPEG compression

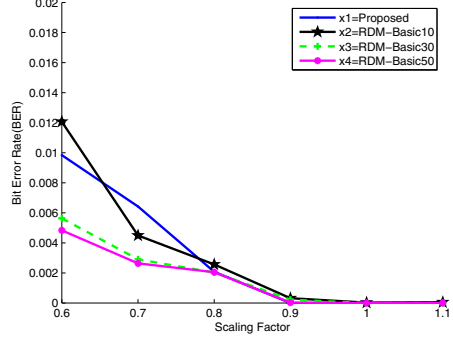


Fig. 5. BER under amplitude scaling attacks

Table 1. BER under gamma correction

Gamma correction	Proposed	RDM-Basic10	RDM-Basic30	RDM-Basic50
$\gamma = 0.9$	0.2160	0.4201	0.5629	0.6271
$\gamma = 1.1$	0.2007	0.4416	0.6086	0.6651

The BER results under image filtering attacks are shown in Table 2. Three filters are used with size of $s \times s$, where $s \in \{3, 5\}$. We can see that the proposed method shows a little better performance than RDM-Basic under the three image filtering attacks.

Additive white Gaussian noise (AWGN) and Salt&Pepper noise are the most commonly used noises in image processing. The watermarked images are distorted by AWGN with standard deviation $\sigma \in \{10, 20\}$ (in the range of $[0, 255]$), and salt&pepper noise with probability $p \in \{0.01, 0.02\}$. The results of the two watermarking methods against noise addition shown in Table 3 demonstrate that the improved RDM method significantly outperforms the basic RDM method under AWGN addition. It is worth noting that the median filter is used before watermark decoding for the Salt&Pepper noise addition attack.

Table 2. BER under image filtering

Image filtering	Proposed	RDM-Basic10	RDM-Basic30	RDM-Basic50
Average filtering (3×3)	0.0159	0.0233	0.0173	0.0150
Average filtering (5×5)	0.1159	0.1484	0.1413	0.1361
Median filtering (3×3)	0.0473	0.0623	0.0650	0.0562
Median filtering (5×5)	0.1363	0.1421	0.1491	0.1309
Wiener filtering (3×3)	0.0018	0.0046	0.0023	0.0019
Wiener filtering (5×5)	0.0435	0.1064	0.0885	0.0843

Geometric attacks always have significant effects on watermarking, while do not cause serious visual distortion of images. Hence, geometric attack is a big challenge for watermarking. Table 4 shows the BER results under rotation and resizing. In our implementation, the rotated images are not rotated back to its original direction. It can be seen that our method performs better than the basic RDM method to resist rotation and resizing attacks.

Table 3. BER under noise addition

Noise addition	Proposed	RDM-Basic10	RDM-Basic30	RDM-Basic50
Gaussian noise ($\sigma = 10$)	0.0323	0.1249	0.0942	0.0804
Gaussian noise ($\sigma = 20$)	0.2401	0.4344	0.4405	0.4400
Salt&Pepper ($p = 0.01$)	0.0512	0.0642	0.0662	0.0579
Salt&Pepper ($p = 0.02$)	0.0562	0.0660	0.0699	0.0608

Table 4. BER under geometric attacks

Attacks	Proposed	RDM-Basic10	RDM-Basic30	RDM-Basic50
Rotation (0.5°)	0.3142	0.3367	0.3667	0.3629
Rotation (-0.5°)	0.3123	0.3511	0.3801	0.3763
Resizing (256×256)	0.0037	0.0079	0.0056	0.0050
Resizing (128×128)	0.0603	0.0814	0.0699	0.0676

Table 5. BER comparison between the proposed method and MWT-EMD [13]

Attacks	Proposed	MWT-EMD	Attacks	Proposed	MWT-EMD
Median filtering (5×5)	0.0046	0.0975	JPEG (Q=10)	0	0
Median filtering (7×7)	0.0353	0.1094	JPEG (Q=20)	0	0
Median filtering (9×9)	0.0896	0.6524	Rotation (1.0°)	0.3250	0.5469
Average filtering (3×3)	0	-	Rotation (0.5°)	0.1009	0.4492
Average filtering (5×5)	0	-	Rotation (-0.5°)	0.1110	0.4414
Gaussian filtering (3×3)	0	0	Rotation (-1.0°)	0.2873	0.5703
Gaussian filtering (5×5)	0	0.0156	Salt&Pepper ($p = 0.08$)	0	0.0284

4.2 Comparison with Other Watermarking Methods

In order to further evaluate the performance of the improved RDM method, we also compare it with two watermarking methods MWT-EMD [13] and GDWM [10]. MWT-EMD is the state-of-the-art method in spread spectrum watermarking and GDWM is one of the state-of-the-art methods in quantization-based watermarking.

Table 5 compares the BER results of the improved RDM method with MWT-EMD method. As in [13], the test images are "Baboon," "Goldhill," "Lena," and "Pepper" and a 64-bit message is embedded in each image with the PSNR of about 42dB. The results of our method are the averaged BERs obtained from embedding 100 different watermarks in each image. It can be seen that our method outperforms MWT-EMD under all the considered attacks.

Table 6 compares the BER results of the improved RDM method with GDWM method. As in [10], the test images are "Baboon," "Barbara," "Lena," and "Pepper" and a 256-bit message is embedded in each image with the PSNR of 43.29dB, 42.70dB, 43.54dB and 43.06dB respectively. We can see that our method is more robust than GDWM in general.

Table 6. BER comparison between the proposed method and GDWM [10]

Attacks	Proposed	GDWM	Attacks	Proposed	GDWM
JPEG (Q=20)	0.0018	0.0154	Rotation (0.5°)	0.2154	0.3715
JPEG (Q=30)	0	0.0034	Rotation (-0.5°)	0.2307	0.3785
JPEG (Q=40)	0	0.0013	Average filtering (3×3)	0	-
Gaussian noise ($\sigma = 10$)	0	0.0146	Average filtering (5×5)	0.0164	-
Gaussian noise ($\sigma = 20$)	0.1433	0.1309	Gaussian filtering (3×3)	0	0
Salt&Pepper ($p = 0.01$)	0.0064	0.0021	Gaussian filtering (5×5)	0.0104	0.0046
Salt&Pepper ($p = 0.02$)	0.0080	0.0088	Median filtering (3×3)	0.0051	0.0182
Salt&Pepper ($p = 0.04$)	0.0199	0.0310	Median filtering (5×5)	0.0613	0.1041

5 Conclusion

In this paper, we have proposed an improved RDM watermarking method. Three aspects are applied to improve the robustness of our algorithm: 1) We increase the quantization step size by modifying two coefficients instead of only one coefficient in the basic RDM method. In this way, the quantization step size is increased. 2) Several modification rules are defined to reduce embedding distortion and to improve robustness. For example, we modify the coefficients according to their magnitude and the relationship between the original ratio and its watermarked ratio. 3) Significant coefficients are selected to embed watermark, because they are more robust and can resist various attacks. A wide range of attacks are tested. Experimental results have verified that our method is not only robust to amplitude scaling attack but also robust to common signal processing attacks. Experiments have also demonstrated that our method has better robustness than the basic RDM and two state-of-the-art watermarking methods, though the capacity of our method is less than that of the basic RDM method. Hence, when considering a robust watermarking, our method is a better choice.

Acknowledgments. The work on this paper was supported by Nature Science Foundation of China (Grant No.61303262) and National Key Technology R&D Program (Grant No.2012BAH04F02).

References

1. Chen, B., Wornell, G.W.: Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Inf. Theory* 47(4), 1423–1443 (2001)
2. Chen, L.H., Lin, J.J.: Mean quantization based image watermarking. *Image and Vision Computing*. 21(8), 717–727 (2003)

3. Lin, W.-H., Wang, Y.-R., Horng, S.-J.: A block-based watermarking method using wavelet coefficient quantization. In: Hua, A., Chang, S.-L. (eds.) ICA3PP 2009. LNCS, vol. 5574, pp. 156–164. Springer, Heidelberg (2009)
4. Chen, T.H., Horng, G., Wang, S.H.: A Robust Wavelet Based Watermarking Scheme using Quantization and Human Visual System Model. *Pakistan Journal of Information and Technology* 2(3), 213–230 (2003)
5. Shterev, I.D., Lagendijk, R.L.: Amplitude scale estimation for quantization-based watermarking. *IEEE Trans. Signal Processing*. 54(11), 4146–4155 (2006)
6. Abrardo, A., Barni, M.: Orthogonal dirty paper coding for informed data hiding. In: Proc. International Society for Optics and Photonics, Electronic Imaging, pp. 274–285 (2004)
7. Bradley, B.A.: Improvement to CDF grounded lattice codes. In: Proc. International Society for Optics and Photonics, Electronic Imaging 2004, pp. 212–223 (2004)
8. Ourique, F., Licks, V., Jordan, R., et al.: Angle QIM: A novel watermark embedding scheme robust against amplitude scaling distortions. In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 2, pp. 797–800 (2005)
9. Zhu, X., Tang, Z.: Improved quantization index modulation watermarking robust against amplitude scaling and constant change distortions. In: Proc. IEEE Int. Conf. Image Processing, pp. 433–436 (2008)
10. Nezhadarya, E., Wang, Z.J., Ward, R.K.: Robust image watermarking based on multiscale gradient direction quantization. *IEEE Transactions on Information Forensics and Security* 6(4), 1200–1213 (2011)
11. Prez-Gonzlez, F., Mosquera, C., Barni, M., et al.: Rational dither modulation: a high-rate data-hiding method invariant to gain attacks. *IEEE Transactions on Signal Processing* 53(10), 3960–3975 (2005)
12. Li, Q., Cox, I.J.: Rational dither modulation watermarking using a perceptual model. In: Proc. IEEE 7th Workshop on Multimedia Signal Processing, pp. 1–4 (2005)
13. Bi, N., Sun, Q., Huang, D., et al.: Robust image watermarking based on multiband wavelets and empirical mode decomposition. *IEEE Transactions on Image Processing* 16(8), 1956–1966 (2007)

The Research of Vehicle Tracking Based on Difference Screening and Coordinate Mapping^{*}

Zhang Jun-yuan¹, Liu Wei-guo^{2,3}, Tong Bao-feng¹, and Wang Nan¹

¹ State Key Laboratory of Automotive Simulation and Control,
Jilin University, Changchun 130022

² Zhejiang Key Laboratory of Automobile Safety Technology,
Hangzhou, 310000
junyuan@jlu.edu.cn

³ GEELY AUTOMOBILE RESEARCH INSTITUTE

Abstract. In order to identify vehicle driving cycle by monocular camera and then offer automotive active safety systems such as ACC (Adaptive Cruise Control) system, accurate condition identification signal, a difference screening method based on haar feature is put forward to identify the vehicle and a method based on coordinate mapping is improved to eliminate the impact that the changes of pitch angle make on the accuracy of positioning the vehicle, then combine with Karman filter technology to track vehicles. Finally, the studied method is used to track the vehicle in an actual video and the test results show that the method can correctly identify the image of vehicle, and accurately track the spatial position of vehicle. As a result, the studied method can be used to offer an active safety system like ACC accurate condition identification signal.

Keywords: Vehicle tracking, Monocular camera, Haar training, Coordinate mapping and Karman filter.

1 Introduction

Vehicle detection technology is one of the important technologies for automotive active safety. Compared with the radar sensor, camera sensor not only is inexpensive, but also can provide richer information. In recent years, along with DSP (Digital Signal Processing) technology developing and the image processing technology maturing, domestic and foreign scholars have made new progress on the research of using camera sensors for vehicle distance measure.

There are two major categories of front vehicle tracking method by using camera: one is based on stereo vision, Young-Chul Lim used sub-pixel accuracy to obtain the front vehicle's position in the image, and then combined with the inverse perspective projection and the kalman filter to track the spatial position and speed of the vehicle ahead [1]. Jonghwan Kim first got the location information, according this information

^{*} Fund: Open Fund of Zhejiang Key Laboratory of Automobile Safety Technology (LHY1109J0565).

to defined the image interest region, and then used haar classifier to identify the vehicle in the region of interest [2]; and Gaojian Cui used the improved region matching tracking algorithm to detect the vehicle [3]. Although the vision method based on stereo is more accurate in positioning the vehicle, there are still some shortcomings like high cost, large computation and low operation speed; the other one is to study the vehicle tracking based on monocular camera. Using a monocular camera for vehicle tracking has been a problem for research. Domestic and foreign scholars have corresponding researches on this field. For instance, Gideon P. Stein used the principle of perspective geometry to do the research of front vehicle locating and speed tracking [4]; Bing-Fei Wu used module matching method to identify vehicles on highways, urban roads and tunnels [5]; and Benrong Wang used the shadow features of the bottom of the vehicle and texture features of images to identify the vehicle, and then located the vehicle via projection transformation [6]. A method based on differential screening and coordinate mapping is proposed in this essay to identify and locate the vehicle. As a result, the accuracy of vehicle identification has been significantly improved and the process of vehicle locating was simplified. Moreover, this method can accurately identify, tracking, positioning the vehicle in a variety of environments and can provide accurate information on working conditions for an active safety equipment like ACC system.

2 Vehicle Identification Based on Goal Differential Screening

Adaboost classifier is used in this paper for vehicle identification. The basic principle of the haar classifier training is that a number of haar features shown in figure 1 are used to threshold the eigenvalues of the rectangular area in images. Some haar features selected to form a weak classifier which allows, in most of the cases, to identify the target area in the image and refuse some non-identification target area. Depending on different haar features, more above mentioned weak classifiers are trained to become strong classifiers, and those classifiers can be used to accurately identify targets. The progress of the target identification is showed in figure 2.

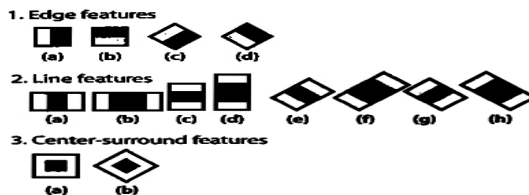


Fig. 1. Haar Features Figure

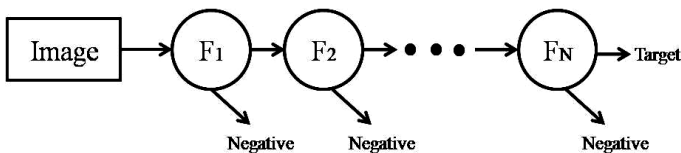


Fig. 2. Process of Target Recognize

In this essay, 11211 samples were collected by a vehicle traveling data recorder, including 3037 positive samples (figure 3) and 8138 positive (figure 4) samples, and then we have trained these samples and obtained a vehicle identification classifier.



Fig. 3. Positive Sample Examples



Fig. 4. Negative Sample Examples

In theory, using haar features adaboost classifier to identify vehicle can achieve high recognition rate, but in the final analysis, the adaboost classifier method for the vehicle identification is achieved based on a series of fixed threshold of haar features. If some areas in the image meet the corresponding threshold requirements of haar features, those areas are identified as vehicles, but the haar eigenvalues of some non-vehicle area in the image can possibly meet the corresponding threshold requirements, thus, in addition to the correctly identified vehicles, there may be some non-vehicle areas are mistakenly identified as the vehicle, as the black circles shown in figure 5, but the error detection is likely random; in contrast, the correctly identified goals change successively between different video frames.

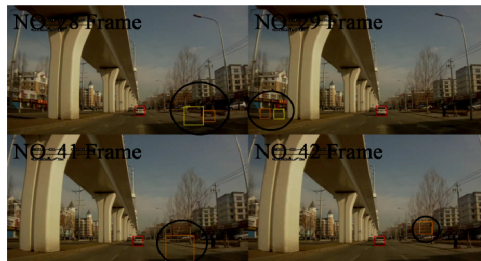


Fig. 5. Result of Vehicle Recognize

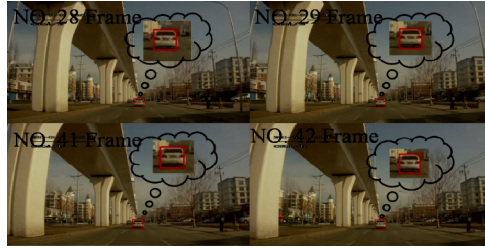


Fig. 6. Result of Vehicle Recognize After Remove err-Recognize

A method for target screening based on the above difference represents the focus of this essay. Assuming that two consecutive image frames respectively identify i and j targets, $P = \{s_1, s_2, \dots, s_i\}$, $Q = \{s_1, s_2, \dots, s_j\}$, both P and Q contain error detection targets and accurate identification targets, each target can be described by k parameters, namely, $s = f(x_1, x_2, \dots, x_k)$. Assuming that there is a smaller difference between the target S_p in P and the target S_q in S , and then S_p and S_q can be considered as the continuously identified same goal, namely,

$$|f(x_{p1}, x_{p2}, \dots, x_{pk}) - f(x_{q1}, x_{q2}, \dots, x_{qk})| \leq e \tag{1}$$

Due to the randomness of false detection, when a target is continuously identified for sufficient times, the target can be considered as correctly identified vehicle. The centroid coordinates (x, y) and the rectangular perimeter L are used to describe a target, if there are two targets between two successive frames which meet the following relationship,

$$\begin{aligned} |x_p - x_q| + |y_p - y_q| &\leq a, \\ |L_p - L_q| &\leq b, \end{aligned} \tag{2}$$

Then these two targets are considered as the same target between the two frames, according to experiences take

$$a = 30, b = 10.$$

In this research paper, the above tracking algorithm is used for video processing. The effects of using this screening method to identify target are shown in figure 6, as can be seen from the figure that only vehicles are screened out, and all the false alarms are removed.

3 Vehicle Tracking and Locating Based on Coordinate Mapping

3.1 Coordinate Mapping Vehicles Ranging

Determining the position of the vehicle in the image is only the initial condition for using camera sensor to provide input information of vehicle driving conditions for the automotive active safety systems such as ACC system. Quantitative description of the

spatial position of the vehicle in front is the automotive active safety system’s real input parameter.

The existing spatial orientation methods by using image information are mostly based on camera calibration, it is not only cumbersome but also for its effects on vehicles ranging is not necessarily sufficient satisfactory. Chi-Feng Wu proposed a method which calculated the mapping relationship between the pixel coordinates and the spatial distance by training neural network [12], however, this method requires the neural network training and this process is cumbersome. In order to simply obtain the spatial position of the vehicle, in this essay is proposed a method that can directly establish a mapping relationship between the portion of the image pixel coordinates and spatial distance.

As shown in figure 7, if the front vehicle is at different distance, in the picture, the vehicle’s horizontal position is also at different. When the camera optical axis is absolutely horizontal, the relationship between the position of front vehicles and pixel coordinates can be obtained from the geometric relationship ,that is,

$$z = \frac{h \cdot f}{y} \tag{3}$$

where, z is the distance of the front vehicle; h is the height of the camera from the ground; f is the focal length of the camera; and y is the vehicle’s horizontal place in the picture.

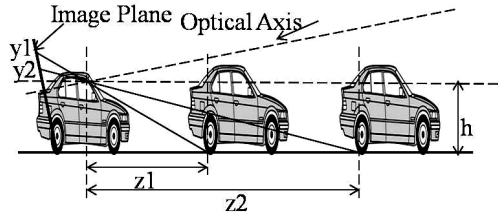


Fig. 7. Camera Imaging System

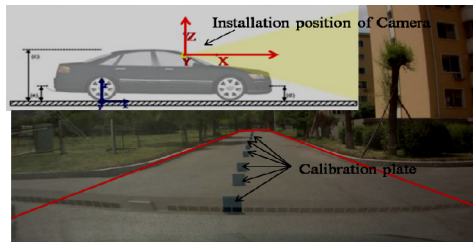


Fig. 8. Schematic View of Coordinate Mapping Experiments

The relationship between the spatial position and the pixel coordinates is theoretically established according to the above relationship, but the absolute horizontal installation of cameras cannot be guaranteed in practical applications, while the pitch angle of camera has great influence on z value[4]. As shown in figure 7, the angle between the camera optical axis and the plane is α ; and the relationship between the

position of front vehicles and pixel coordinates can be obtained from the geometric relationship ,that is,

$$z = \frac{f(z \tan \alpha + h) \cos^2 \alpha}{y} + \frac{(z \tan \alpha + h) \cos \alpha \sin \alpha}{y} \tag{4}$$

The following method is used to establish the relationship between the position of front vehicles and pixel coordinates.

First, the camera is fixed in front and rear of the rearview mirror in the vehicle, and some calibration boards are put at some places with known spatial distance in front of the vehicle in a reasonable flat and wide space, as shown in figure 8; then record image information; and finally, get the pixel ordinate at the lower edge of the calibration board and fit the curve about the relationship between pixel ordinate and spatial distance. This fitted curve can be regarded as the mapping relationship of pixel ordinate and spatial distance.

In this essay, the camera sensor is installed on a passenger car, and the calibration boards are placed in front of the camera sensor, propose corresponding function model, fit curve by using the least squares method and repeatedly amend it. Eventually, the fitting result is,

$$\ln(z) = 0.026xe^{0.995x} - \frac{1.845}{53.58x - 43.47} + 1.708 \tag{5}$$

where $x = (a \times y) / 100$, a is the height of pixel point.

The fitting curve, as shown in figure 9, describes the mapping relationship between the spatial distance and the ordinate of pixel, and the corresponding spatial distance of pixel can be obtained from the curve; therefore, this can be used to locate the front vehicle.

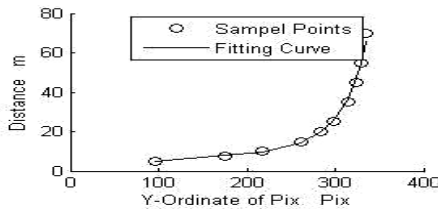


Fig. 9. Fitting Curve of Pixel Ordinate and Distance



Fig. 10. Simulation Scenarios of PreScan

In order to verify the feasibility of the method, it is validated in the PreScan¹ simulation environment.

The host vehicle following the front vehicle, obtain the bottom (Red line position in figure 10) of front vehicle's ordinates pixels, calculate the spatial position of vehicle by using the relation formula (5), compare this position information with the spatial position of the front vehicle detected by the PreScan's ideal radar sensor, the compared result is shown in figure 11, as can be seen in figure 11, it can meet the accuracy requirements of vehicles ranging.

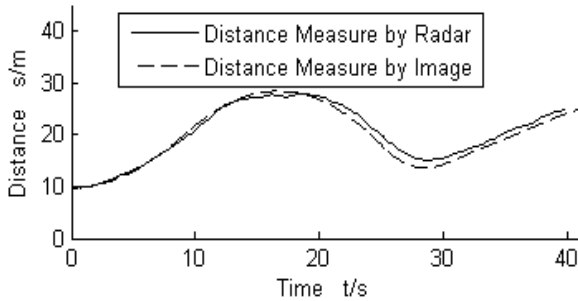


Fig. 11. Distance Measurement by Camera and Radar

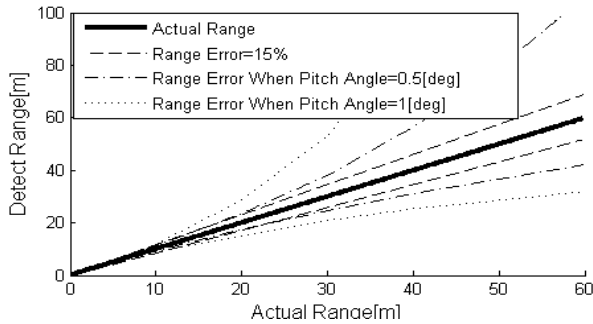


Fig. 12. Distance Measurement Influence by Pitch Angle

3.2 The Study of Elimination the Impact of Pitch Angle Ranging

When the vehicle runs at the bumpy road or force break which would cause the change of the pitch angle of the camera sensor, the distance measure accuracy will be seriously affected. Research shows that when the pitch angle changes $\pm 1^\circ$, ranging

¹ PreScan is an automotive active safety research software which was developed by TNO company. It can be used to build a very convenient scenario, and it also has many sensor models (such as radar, camera and radio) which can be used to detect the vehicle driving information. Additionally, PreScan has a very convenient interface with Simulink, so it can be used to do research on the automotive active safety by means of Simulink. PreScan also has a 3D viewer window, which can be used to observe the vehicle's driving condition.

error will be 15% greater, as shown in figure 12. While when the vehicle is in emergency conditions, and the driver breaks the vehicle, the pitch angle will reach $\pm 2.5^\circ$; therefore, it is necessary to study methods to eliminate the affects made by pitch angle changes on the accuracy of measurement.

After the study of imaging system, we have realized that when the pitch angle of vehicle is 0° , the space plane with the same height as camera sensor will be a line after imaging, and no matter how far the front vehicle is, the vehicle in the image is divided into two parts by this line and the ratio of the pixel height of the two parts is constant. In this essay, this line is defined as fixed ratio distribution line as shown in figure 13.

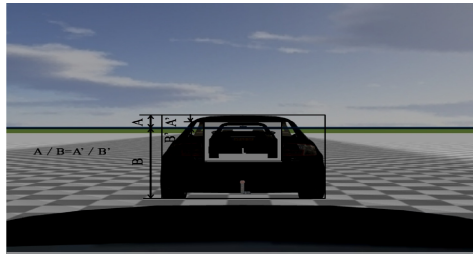


Fig. 13. Schematic View of Features of Contour plane

Providing that the fixed ratio distribution line $f(x)=y_0$ divides the vehicle into two parts, the ratio of pixel height is then r . When the vehicle pitch angle changes during the running, the vehicle in the image fluctuates as shown in figure 14. Provided that at a certain moment the ratio of the divided parts by $f(x)=y_0'$ is r , then because of the changes of the pitch angle, the vehicle in the image will longitudinally fluctuate $\Delta y=y_0'-y_0$ pixels. If at this time the coordinate of pixel at the bottom of the vehicle in the image is y_b' , then we can figure out that at the current distance from vehicle; if the pitch angle of vehicle is 0° , the ordinate of the bottom of the vehicle in the image is,

$$y_b = y_b' - \Delta y = y_b' - (y_0' - y_0) \tag{6}$$

The value of r must be completely calculated at the initial stage of the detection of vehicle. Assuming that at the beginning of the detection of the vehicle, the camera sensor keeps horizontal, and it continuously extract N frame images and calculate the average value of the ratio r of these images.

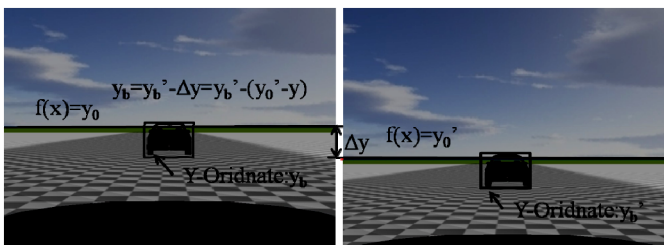


Fig. 14. Schematic View of Influence by Pitch Angle Change

When the pitch angle of vehicle changes according to the characteristic of the fixed ratio distribution line and the value of r , we can calculate the value of Δy , and use the formula (6) to calculate y_b , then the actual spatial distance of front vehicle corresponding to y_b can be calculated according to formula (5). Thus the effects made by the changes of pitch angle on distance measure can be eliminated.

PreScan is used to build simulation scenario to verify the correctness of the method in this research paper. When the vehicle runs on bumpy road, radar and camera sensor are used to detect the distance from the front vehicle as shown in figure 15. As one can see from the figure, the effects made by the changes of pitch angle on ranging can be resisted by this method.

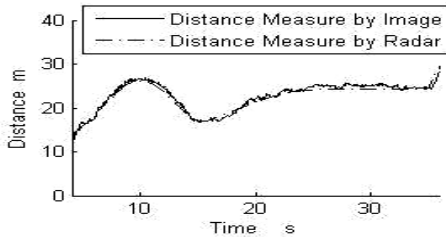


Fig. 15. Distance Measurement Result of Eliminate the Influence of Pitch Angel Change

3.3 Vehicle Tracking by Kalman filtering

The vehicle identification used by classifier (trained by adaboost algorithm) can only obtain an approximate area of the vehicle in the image, rather than accurate to each pixel. Therefore when the vehicle is directly located by this mapping relationship, a phenomenon will appear: the calculated distance from the front vehicle violently fluctuates (the thin blue line shown in figure 16). To avoid the impact made by the above phenomenon and guarantee the wanted impact on vehicle location, kalman filter technology is used for vehicle tracking in this essay. In order to achieve the kalman filter, mathematical model of the target movement is built first. A dynamic model of discrete control system is introduced in the following way:

$$X'(t) = AX(t) + \omega(t) \quad (7)$$

In additional to the system dynamic model, the system measurement equation is also introduced to describe the relationship between the measured value and the target moving state as shown in the following formula:

$$Z(t) = HX(t) + v(t) \quad (8)$$

The above kalman filter method is used to filter the distance from the front vehicle. The filtered and unfiltered vehicle distances curve are shown in figure 16. As one can see on figure 16, the filtered vehicle distance is relatively smooth and stable.

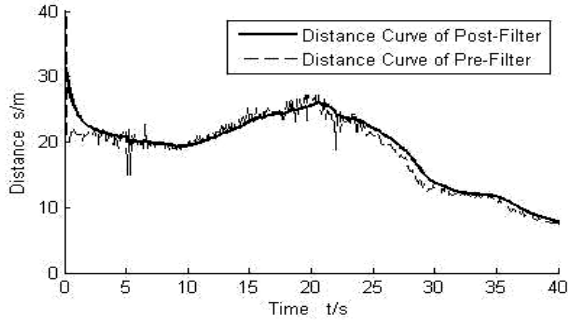


Fig. 16. Result of Karman Filter of the Distance

4 Experiments

In order to verify the performance of the practical application of the proposed method, camera (controlled by a C++program?) is used to captured the video of the actual driving. The vehicles in these videos are identified and located by using goal differential screening and coordinates mapping because the 1280×720 images on the PC can handle about 10 frames per second. The results are shown in figure 17: there are totaly 396 frames in this segment of video, and according to the statistics, there are in total 25 frames with detection errors; the error rate is 6.3%. The errors are mainly due to the missing of detection, but the current distance between vehicles can still be estimated by kalman filter when the vehicles are undetected (the results of vehicle locating shown in figure 17). Thus, it will not have much impact on the locating of the



Fig. 17. Result of Vehicle Tracking and Location

vehicle. As a result, the proposed method has been approved with its satisfying performance in identifying and locating vehicle in actual driving situation.

5 Conclusion

A method of vehicle tracking and distance measure is supposed in this essay, experimental results show that this method is reasonable and reliable; and the results of vehicle tracking is satisfactory. The studied method in this research paper can provide automotive active safety systems such as ACC system as well as driving condition information.

The ability of the identification can be improved by expanding the number of positive and negative samples and increasing the variety of environmental samples (rainy, night, etc.). Therefore, further research will enrich sample library. In the future, the location of vehicle on the curve line will be studied with the combination of lane recognition technology.

References

1. Lim, Y.-C., Lee, M., Lee, C.-H., et al.: Improvement of Stereo Vision-based Position and Velocity Estimation and Tracking Using a Stripe-based Disparity Estimation and Inverse Perspective Map-based Extended Kalmanfilter. *Optics and Lasers in Engineering* 48, 859–868 (2010)
2. Kim, J., Lee, C.-H., Lim, Y.-C., Kwon, S.: Stereo Vision-Based Improving Cascade Classifier Learning for Vehicle Detection. In: *Bebis, G., et al. (eds.) ISVC 2011, Part II. LNCS, vol. 6939, pp. 387–397. Springer, Heidelberg (2011)*
3. Cui, G., Huang, Y., Tian, Y.: Detection for Forward Vehicle Based on Stereo Machine Vision. *Computer Automated Measurement & Control* 13(9), 890–892 (2005)
4. Stein, G.P., Mano, O., Shashua, A.: Vision-based ACC with a Single Camera: Bounds on Range and Range Rate Accuracy. In: *IEEE Proceedings of the Intelligent Vehicles Symposium, June 9-11, pp. 120–125 (2003)*
5. Wu, B.-F., Kao, C.-C., Liu, C.-C., et al.: The vision-based vehicle detection and incident detection system in Hsueh-Shan tunnel. In: *IEEE International Symposium on Industrial Electronics, ISIE 2008, June 30-July 2, pp. 1394–1399 (2008)*
6. Rongben, W., Boyuan, G., Lie, G., et al.: A Study on Multiple Vehicle Detection Based on Computer Vision. *Automotive Engineering* 28(10), 902–905 (2006)
7. Shimomura, N., Fujimoto, K., Oki, T., et al.: An Algorithm for Distinguishing the Types of Objectson the Road Using Laser Radar and Vision. *IEEE Transactions on Intelligent Transportation Systems* 3(3), 189–195 (2002)
8. Dingcong, P.: Basic Principle and Application of Kalman Filter. *Software Guide* 8(11) (November 2009)
9. Miyahara, S., Sielagoski, J., Koulinitch, A.: Target Tracking by a Single Camera Based on Range-Window Algorithm and Pattern Matching. In: *2006 SAE World Congress Detroit, Michigan, April 3-6 (2006)*
10. Qing, X., Feng, G., Guoyan, X.: An Algorithm for Front-vehicle Detection Based on Haar-like Feature. *Automotive Engineering* 35(4) (2013)

11. Mai, X., Yang, M., Wang, C., et al.: Multi sensor Fusion Based Vehicle Detection and Tracking Method. *Journal of Shanghai Jiaotong University* 45(7) (July 2011)
12. Wu, C.-F., Lin, C.-J., Lee, C.-Y.: Applying a Functional Neurofuzzy Network to Real-Time Lane Detection and Front-Vehicle Distance Measurement. *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews* 42(4), 577–589 (2012)

Pathology Image Retrieval by Block LBP Based pLSA Model with Low-Rank and Sparse Matrix Decomposition

Yushan Zheng, Zhiguo Jiang, Jun Shi, and Yibing Ma

Image Processing Center, School of Astronautics, Beihang University
Beijing Key Laboratory of Digital Media
Beijing, China

yszheng@sa.buaa.edu.cn jiangzg@buaa.edu.cn
chris.shi331@gmail.com hemp110@126.com

Abstract. Content-based image retrieval (CBIR) is widely used in Computer Aided Diagnosis (CAD) systems which can aid pathologist to make reasonable decision by querying the slides with diagnostic information from the digital pathology slide database. In this paper, we propose a novel pathology image retrieval method for breast cancer. It firstly applies block Local Binary Pattern (LBP) features to describe the spatial texture property of pathology image, and then use them to construct the probabilistic latent semantic analysis (pLSA) model which generally takes advantage of visual words to mine the topic-level representation of image and thus reveals the high-level semantics. Different from conventional pLSA model, we employ low-rank and sparse matrix composition for describing the correlated and specific characteristics of visual words. Therefore, the more discriminative topic-level representation corresponding to each pathology image can be obtained. Experimental results on the digital pathology image database for breast cancer demonstrate the feasibility and effectiveness of our method.

Keywords: Image retrieval, computer aided diagnosis, breast cancer, probabilistic latent semantic analysis, low-rank and sparse matrix composition.

1 Introduction

Computer Aided Diagnosis (CAD) system for breast cancer has attracted more and more attention due to morbidity increase of breast cancer in female [1, 2]. Although new technologies for breast cancer diagnosis have developed rapidly in the past few years, the final diagnosis still relies on the pathological theories [3]. As the digital pathology slides spread, senior pathologists can mark the lesion area with detailed descriptions on the digital slides and share it to others through CAD systems or the Internet. In the other hand, junior pathological doctors can get valuable suggestions by searching slides that contain diagnosis information when facing indeterminable cases. Therefore, CAD systems consisting of pathology slide database with confirmed diagnosis information are urgently required. But it is challenging to retrieve useful slides from the enormous database effectively and accurately for the reason that the resolution of digital pathology

slide is usually much higher than common digital image and the characteristics of pathology image are much different from natural images.

To deal with the retrieval problem on digital pathology slide databases, Content-Based Image Retrieval (CBIR) has been proposed and successfully applied to clinical diagnosis [4, 5]. Over the past years, a large number of retrieval methods for pathology image have been proposed. Caicedo et al. [6] apply different kinds of visual features to achieve the retrieval task for four kinds of tissues. Kowal et al. [7] take advantage of statistical features of individual nuclei to classify benign and malignant cases of breast cancer. However, these methods mentioned above just describe the global characteristics of the image and may even ignore the high-level semantics that exist in the image. Therefore, to mine the texture information and local property of pathology image, we propose to divide the entire image into non-overlapping blocks and extract Local Binary Pattern (LBP) [8] descriptor in each block. Then LBP descriptors are used to build the codebook composed of visual words through k-means. Afterwards, each pathology image can be represented by the word frequency histogram via Bag-of-Words (BoW) [12] scheme. However, there are synonyms among visual words and thus make the word-level representation hard to discriminatively reveal the semantics in images. Therefore, probabilistic latent semantic analysis (pLSA) [9] model is applied in our method to mine the high-level semantics of words. Nonetheless, pLSA model just uses BoW scheme to discover the word distribution, which is likely to ignore that there are some correlated and specific characteristics among words. Consequently, the word-level representation of conventional pLSA model may fail to describe the image content precisely. To improve the discriminant ability of pLSA, we apply low-rank and sparse matrix decomposition technique [10, 13] to decompose the word-level representation into two meaningful parts (i.e., correlated and specific word-level representations), and then utilize them to train two pLSA models. Finally each image can be represented by the combination topics learned from these two models.

Our proposed method consists of two contributions. First, we use block LBP features to describe the spatial texture information and then apply pLSA model to discover the high-level semantics of pathology images. The second is that we use the low-rank and sparse matrix decomposition to obtain two word-level representations which can characterize the correlated and specific parts of the visual word distribution. As a consequence, the discriminant ability of word-level representation has been greatly improved. Experimental results on the digital pathology image database of breast cancer demonstrate the feasibility and effectiveness of our method.

The rest of the paper is arranged as follows: Section 2 introduces block LBP descriptor. Section 3 describes the pLSA model along with the low-rank and sparse matrix decomposition. Section 4 presents the experimental database and results. Finally the conclusion is given in Section 5.

2 Block Local Binary Pattern (LBP)

Local binary Pattern (LBP) [8] is a powerful local texture descriptor with the advantages of rotation invariance and orientation invariance. The pattern of each pixel is

calculated by quantifying pixels of its neighborhood into a string of binary code. Generally, the size of neighborhood is defined to 3×3 . The basic LBP code of the central pixel is computed as

$$LBP(p_c) = \sum_{i=0}^7 2^i b(g(p_i) - g(p_c)) \tag{1}$$

where p_c is the central pixel and p_i is the neighbor pixel of p_c , $g(p)$ is the gray value of p and $b(u)$ is the binary function that $b(u)=1$ if u is greater or equal to 0; $b(u)=0$ otherwise. The process of LBP is given in Fig. 1.

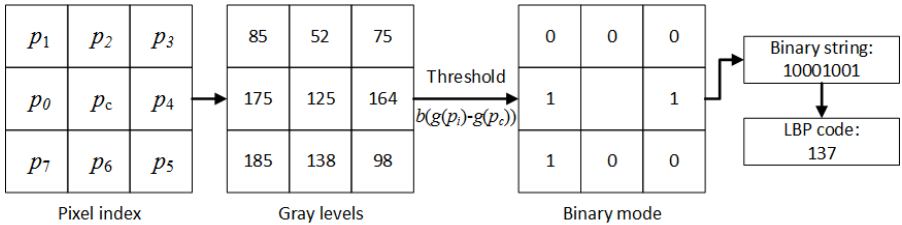


Fig. 1. The process of LBP

It can be found that the image is represented by a 256-dimensional (8-bits binary codes stand for 256 numbers in total) histogram by counting the pixel number of each LBP code. However, the 256-dimensional representation is redundancy and only 58 codes reflect the primitive structural information such as edges and corners. Then the dimension of the histogram is usually reduced to 59 by assigning non-uniform patterns to single bin [11].

It is obvious that LBP histogram or its uniform pattern is a global texture descriptor. To further discover the local structures, we divide the entire image into the non-overlapping blocks (16×16) and then compute uniform pattern of LBP in each block. Finally, pathology images can be represented by a sequence of 59-dimensional LBP histograms.

3 High-Level Semantic Mining

3.1 Probabilistic Latent Semantic Analysis (pLSA)

Although block LBP features mentioned above can characterize the pathology images, they are likely to ignore the high-level semantics that may exist in the image. As the high-level semantic model, Bag-of-words (BoW) [12] performs k -means clustering on the local feature descriptors to generate the codebook composed of visual words, and then quantizes the descriptors into the words through nearest neighbor principal. Finally the image can be represented by words. However, there are synonyms among visual words, which may cause that the semantics of images are not well revealed. As a well-known topic model, probabilistic latent semantic analysis (pLSA) [9] model aims to describe the image content by the latent topic-level representation

learned from the visual words. Moreover, it has simplicity and low computational complexity.

Let $\mathbf{Z} = [z_1, \dots, z_T]$ be the set of latent topics between documents $\mathbf{D} = [d_1, d_2, \dots, d_M]$ and words $\mathbf{W} = [w_1, w_2, \dots, w_N]$. The goal of pLSA is to learn the latent topic probability distribution through the joint probability distribution of documents \mathbf{D} and words \mathbf{W} . Specifically, for image retrieval application, \mathbf{D} is a dataset of images, and \mathbf{W} is the collection of visual word representations in the dataset and \mathbf{Z} can be viewed as the latent variables between \mathbf{W} and \mathbf{D} , namely the topic-level representation. The graph model representation of pLSA is shown in Fig. 2.

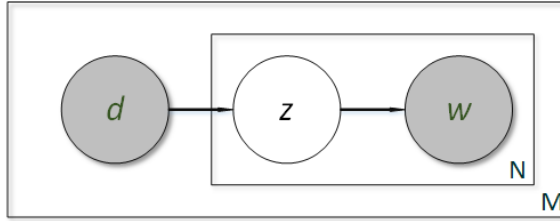


Fig. 2. Graph model representation for pLSA

The joint probability between \mathbf{W} and \mathbf{D} is defined by Eq. (2):

$$P(d_i, w_j) = P(d_i)P(w_j | d_i), \quad P(w_j | d_i) = \sum_{k=1}^T P(z_k | d_i)P(w_j | z_k) \quad (2)$$

where $P(d_i)$ denotes the probability which d_i occurs, $P(z_k | d_i)$ is probability distribution of latent topic z_k in document d_i and $P(w_j | z_k)$ is the probability distribution of topic z_k on word w_j . pLSA model can be viewed as a maximum log-likelihood formulation:

$$\begin{aligned} L &= \sum_{i=1}^N \sum_{j=1}^M n(d_i, w_j) \log P(d_i, w_j) \\ &= \sum_{i=1}^N n(d_i) \left[\log P(d_i) + \sum_{j=1}^M \frac{n(d_i, w_j)}{n(d_i)} \log \sum_{k=1}^T P(w_j | z_k) P(z_k | d_i) \right] \end{aligned} \quad (3)$$

where $n(d_i, w_j)$ represents the frequency that word w_j occurs in document d_i and $n(d_i)$ denotes the occurrence frequency of d_i . Therefore, the solution of pLSA model is to seek the optimal $P(z_k | d_i)$ and $P(w_j | z_k)$ through expectation-maximization (EM) algorithm [9], and $P(z | d_i)$ is the topic-level representation of the i -th document.

3.2 Low-Rank and Sparse Matrix Decomposition

According to [10], the word-level representation generated by BoW implies both correlated and specific information, and each of these two parts is more robust and discriminative for representing the image content. In this paper, we apply the low-rank and sparse matrix decomposition method to decompose the BoW representation (*i.e.*, the word-level representation) of the images into two parts (*i.e.*, low-rank part and sparse part). After decomposition, the low-rank part can reveal the correlated

characteristics of words and the sparse part can indicate the specific characteristics of words. In other words, we obtain two word-level distributions that can describe the generality and speciality of words through low-rank and sparse matrix decomposition.

As mentioned above in Section 3.1, $\mathbf{W} = [w_1, w_2, \dots, w_N]$ is the collection of BoW representations where w_i is the representation of i -th training image. Thus the decomposition is defined as:

$$\mathbf{W} = \mathbf{L} + \mathbf{S} \quad (4)$$

where \mathbf{L} and \mathbf{N} are the low-rank matrix and the sparse matrix. The problem of low-rank and sparse matrix decomposition can be characterized by

$$\begin{aligned} \min_{L,N} \text{rank}(\mathbf{L}) + \lambda \|\mathbf{S}\|_0 \\ \text{s.t. } \mathbf{W} = \mathbf{L} + \mathbf{S} \end{aligned} \quad (5)$$

The $\|\cdot\|_0$ is zero-norm that counts the non-zero elements in the matrix and $\lambda > 0$ is the coefficient that balances the rank term and the sparsity term. Since the problem is non-convex and hard to solve, it can be approximated by solving (6) according to [13]:

$$\begin{aligned} \min_{L,N} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \\ \text{s.t. } \mathbf{W} = \mathbf{L} + \mathbf{S} \end{aligned} \quad (6)$$

The $\|\cdot\|_*$ is the nuclear norm defined as the sum of all singular values. The problem can be solved by the augmented Lagrange multiplier method (ALM) proposed by Lin et al [14].

3.3 Low-Rank and Sparse Matrix Decomposition Based pLSA Model

After the low-rank and sparse matrix decomposition, we obtain a low-rank matrix \mathbf{L} which can characterize the correlated part of words and a sparse matrix \mathbf{N} which can characterize the specific part of words. Each column vector l_i of the matrix $\mathbf{L} = [l_1, l_2, \dots, l_N]$ can be regarded as the representation of correlated characteristics of the i -th training image, and each column vector n_i of the matrix $\mathbf{S} = [s_1, s_2, \dots, s_N]$ implies the specific characteristics. Therefore, instead of w_i , we respectively apply l_i and n_i for the word-level representations of i -th image, and then use them to train two pLSA models. Note that l_i and n_i are L_1 -normalized after absolute operation.

The flow chart of our work is presented in Fig. 3. First we extract the 59-dimensional block LBP histogram for each pathology images in the training set. Then the codebook can be gained through k -means and the word-level representation (namely w_i in matrix \mathbf{W}) corresponding to each image is quantized. After the low-rank and sparse matrix decomposition step, the matrices \mathbf{L} and \mathbf{S} take place of \mathbf{W} to be the word-level representations. EM algorithm is respectively used to compute the optimal $P(z|d)$ and $P(w|z)$ of these two representations, and the combination of $P(z|d)$ is the final topic-level representation of each image. In the test stage, the input ROI

will be converted to the topic-level representation learned from correlated and specific word-level distributions. After computing the similarities between ROI image and the images stored in the database, the top R similar images along with the confirmed diagnosis information are returned to the CAD system. By comparing ROI with these returned images, pathologists can make a more reliable diagnosis decision.

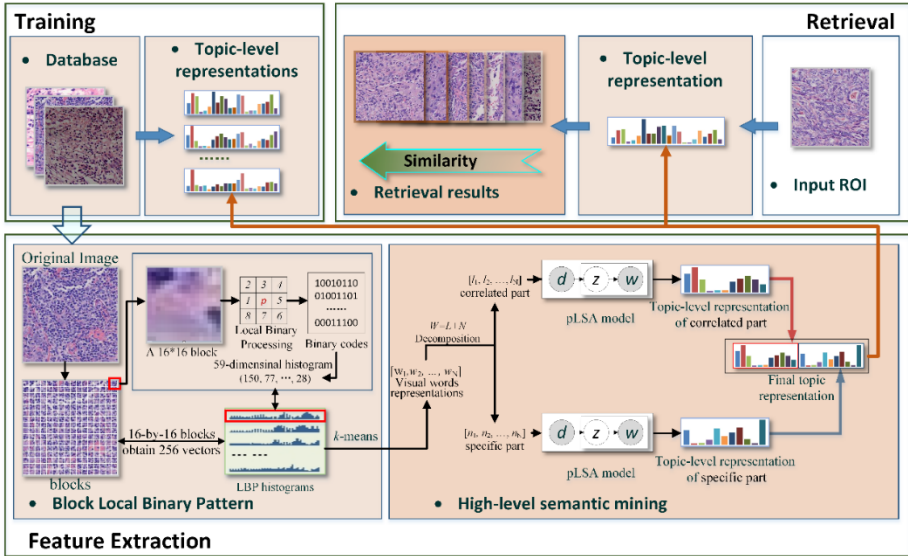


Fig. 3. The flow chart of our retrieval framework

4 Experiment

Our proposed method is evaluated on the pathology image database for breast cancer with confirmed diagnosis information, which is from Motic digital slide database for the yellow race [20]. The image database consists of 5 categories and 600 images (256x256, 20x magnification) for each category, as shown in Fig. 4.

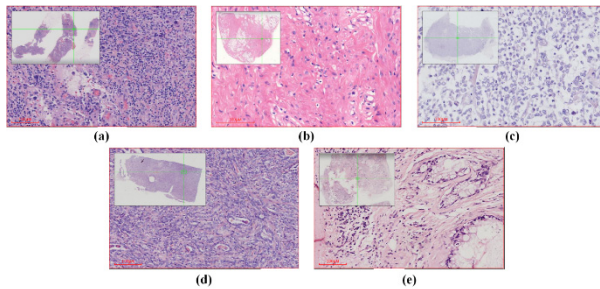


Fig. 4. Five categories of digital pathology slides. (a) Basal-like carcinoma (BLC). (b) Breast myofibroblastoma (BMFB). (c) Invasive breast cancer (IBC). (d) Low-grade adenosquamous carcinoma (LGASC). (e) Mucinous cystadenocarcinoma (MCA).

To evaluate the performance of our method, we compare it with block-LBP-based BoW (LPB-BoW), block-LBP-based pLSA (LBP- pLSA), along with approaches of Caicedo et al. [6] and Kowal et al [7]. Note that we perform 20 times to randomly select 300 images of each category for training and the remaining for test. For each time, we calculate the mean Average Precision (mAP) these five methods for top 20 returned images through cosine distance based similarity measure. Table 1 shows the performances of these five methods.

Table 1. Performance comparison at the top 20 returns of five methods

Algorithm	Performance
Kowal et al. [7]	53.1±1.61
Caicedo et al. [6]	70.2±1.06
LPB-BoW	74.4±1.17
LBP-pLSA	76.3±0.41
Our method	78.6±0.49

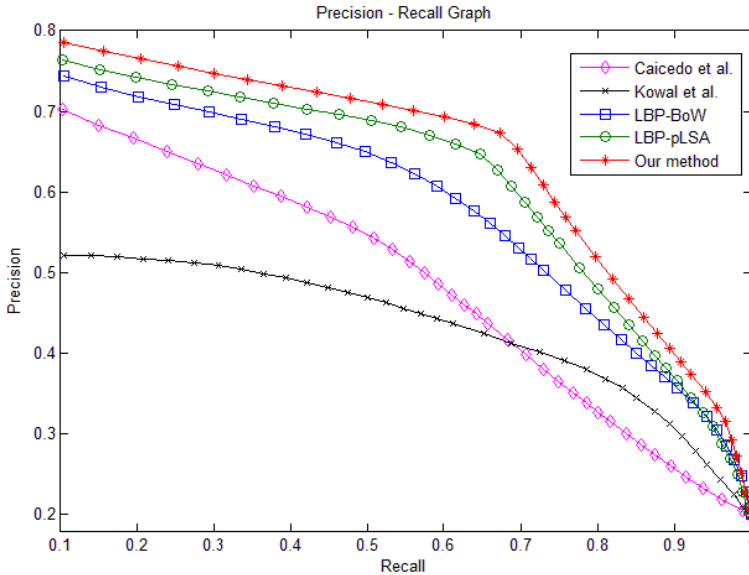


Fig. 5. Precision-recall curves of five methods

As can be seen from Table 1, compared with the methods proposed by Caicedo et al. [6] and Kowal et al [7], LPB-BoW has superior retrieval performance. It may be due to the fact that block-LBP features can effectively describe the spatial property of texture structure, and in the other hand, it may benefit from the semantic characterization ability of BoW. Particularly, as pLSA model overcomes the limitation of BoW, LBP-pLSA and our method are better. It should be noted that our method is more excellent than LBP-pLSA, since it can discover the correlated and

specific parts of the visual word distribution which leads to the more discriminant word-level representation. The precision-recall curves of these methods are presented in Fig. 5. It indicates that our method overall outperforms the others.

5 Conclusion

In this paper, we propose a novel pathology image retrieval method for breast cancer. Block LBP descriptor is used to describe the spatial characteristics of texture structure. Then they are applied to generate into visual word representation by BoW scheme. After low-rank and sparse composition operating, the word-level representation of each image is decomposed into correlated part and specific part. Based on these two parts, two pLSA models are learnt to mine the high-level semantics existed in the images. Finally, each image is represented by the combined topics of the two pLSA models. Experiments on the pathology image database for breast cancer demonstrate the effectiveness of our method. Further research will aim to apply Local Sensitive Hashing (LSH) to boost the efficiency of retrieval when facing large database.

Acknowledgement. This work was supported by the National Natural Science Foundation of China (No. 61371134), and the 973 Program of China (Project No. 2010CB327900).

References

1. Rebecca, S., Deepa, N., Ahmedin, J.: Cancer Statistics. *CA Cancer Journal for Clinicians* 63(1), 11–30 (2013)
2. Li, N., Zheng, R.S., Zhang, S.W., Zou, X.N., Zeng, H.M., Dai, Z., Chen, W.Q.: Analysis and Prediction of Breast Cancer Incidence Trend in China. *Chinese Journal of Preventive Medicine* 46(8), 703–707 (2012)
3. Fu, X.L.: *The Atlas for Pathologic Diagnosis of Breast Tumours*. Scientifics and Technical Documents Publishing House, Beijing (2013)
4. Xue, Z.Y., Long, L.R., Antani, S., Thoma, G.R.: Pathological-based Vertebral Image Retrieval. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, Chicago, pp. 1893–1896 (2011)
5. Lijia, Z., Shaomin, Z., Dazhe, Z., Hong, Z., Shukuan, L.: Medical Image Retrieval Using Sift Feature. In: *IEEE 2nd International Congress on Image and Signal Processing*, pp. 1–4. Tianjin (2009)
6. Caicedo, J.C., Izquierdo, E.: Combining Low-level Features for Improved Classification and Retrieval of Histology Images. *Transactions on Mass-Data Analysis of Images and Signals* 2(1), 68–82 (2010)
7. Marek, K., Paweł, F., Andrzej, O., Józef, K., Roman, M.: Computer-aided Diagnosis of Breast Cancer Based on Fine Needle Biopsy Microscopic Images. *Computers in Biology and Medicine* 43(10), 1563–1572 (2013)
8. Ojala, T., Pietikäinen, M., Harwood, D.: A Comparative Study of Texture Measures with Classification Based on Featured Distributions. *Pattern Recognition* 29(1), 51–59 (1996)

9. Hofmann, T.: Probabilistic Latent Semantic Analysis. In: The 22nd Annual ACM Conference on Research and Development in Information Retrieval, San Francisco, pp. 289–296 (1999)
10. Zhang, C., Liu, J., Tian, Q., et al.: Image Classification by Non-Negative Sparse Coding, Low-Rank and Sparse Decomposition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1673–1680 (2011)
11. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(7), 971–987 (2002)
12. Fei-Fei, L., Perona, P.: A Bayesian Hierarchical Model for Learning Natural Scene Categories. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, pp. 524–531 (2005)
13. Candes, E., Li, X., Ma, Y., Wright, J.: Robust Principal Component Analysis? *Journal of the ACM* 58(3), 11 (2011) (submitted)
14. Lin, Z., Chen, M., Ma, Y.: The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices. *arXiv preprint arXiv, 1009.5055* (2010)
15. Motic digital slide database,
<http://med.motic.com/SlideLibraryList.aspx>

Remote Sensing Image Change Detection Based on Low-Rank Representation^{*}

Yan Cheng^{1,2}, Zhiguo Jiang^{1,2}, Jun Shi^{1,2}, Haopeng Zhang^{1,2}, and Gang Meng³

¹ Image Processing Center, School of Astronautics, Beihang University, Beijing, China

² Beijing Key Laboratory of Digital Media, Beijing, China

{cy36152112, buaazhp}@126.com, jiangzg@buaa.edu.cn,
chris.shi331@gmail.com

³ Beijing Institute of Remote Sensing Information, Beijing, China

menggangmark@126.com

Abstract. In this paper we propose an unsupervised approach based on low-rank representation (LRR) for change detection in remote sensing images. Given a pair of remote sensing images obtained from the same area but in different time, the subtraction and logarithm ratio operators are firstly applied to obtain two difference images. Meanwhile the sparse part generated by LRR is also employed for acquiring another difference image, which can detect the change information. Afterwards, LRR is used again to obtain the low-rank part of these three difference images which can reflect the common characteristics. Finally k -means is performed on the low-rank part and thus the final result of change detection can be gained. Experimental results show the effectiveness and feasibility of the proposed method.

Keywords: Change detection, Remote sensing, Low-rank representation, K-means.

1 Introduction

Change detection plays a crucial role in the analysis and understanding of multi-temporal remote sensing images, with wide applications in both civil and military domains, such as agricultural survey [1], forest monitoring [2], natural disaster monitoring [3], urban change analysis [4], etc.

Change detection is the process of identifying differences in the state of an object or phenomenon by observing it at different times [5]. This problem has been given more attention in the past decades. Change detection algorithms are broadly divided into two categories: the supervised and unsupervised approaches [1, 5]. The supervised approaches usually need multi-temporal ground truth for training. However, such ground truth is difficult to obtain in practical applications. The unsupervised approaches obtain the comparison results from the remote sensing images directly, such as image differencing, change vector analysis (CVA) [6], principal component

* The project is supported by National Natural Science Foundation of China (61074029) and Natural Science Foundation of Liaoning Province (20102014)

analysis (PCA) [7] and vegetation index differencing [8]. Therefore, in this paper, we focus on the problem of unsupervised change detection.

As mentioned in [5], the procedure of unsupervised change detection algorithms has three major steps generally: (1) image preprocessing, (2) obtaining the difference image, and (3) analyzing the difference image and post-processing. Among them, the construction of difference map can greatly affect the result of change detection. Although we can use some conventional methods (*e.g.*, subtraction or logarithm ratio operator), they cannot make full use of multi-temporal remote sensing images. Therefore, how to obtain more accurate difference image is still an open problem.

On the other hand, low-rank representation (LRR) [9] has attracted wide concern in computer vision and machine learning field. Compared with sparse representation (SR) [10, 11] which computes the sparse representation of each data vector individually, LRR makes use of matrix decomposition to obtain the low-rank and sparse parts of data vectors jointly which can reveal the common and specific characteristics of data. Considering the problem of change detection, if we use LRR to decompose the matrix consisting of multi-temporal images, the low-rank part will correspond to the unchanged areas and the sparse part will correspond to the changed areas. In other words, the difference image can be generated by the sparse part. Therefore, LRR can effectively detect the change information from a global perspective.

Motivated by the above discussion, in this paper, we propose an unsupervised method for change detection, which applies subtraction operator, logarithm ratio operator and LRR to construct the difference image. Specifically, we firstly use the subtraction and logarithm ratio operators to generate two difference images. At the same time, we apply LRR to decompose the data matrix which is composed of the image before change and the image after change. Based on these three difference images, LRR is used again to extract the low-rank part from three difference images, which can reflect the common characteristics of them and thus can be viewed as the final difference result. Finally k -means is applied to cluster the final difference image into two clusters. Experiments on real remote sensing images demonstrate the feasibility and effectiveness of the proposed method.

The contribution of this paper includes two aspects. On one hand, we propose a novel method of generating the difference image through LRR. Since LRR can describe the common and specific characteristics of data, it can detect the change information of multi-temporal images. On the other hand, we use the LRR-based fusion scheme to combine multiple difference images constructed by various change detection methods, which can effectively improve the stability of change detection.

The remainder of this paper is organized as follows: Section 2 describes low-rank representation (LRR) in detail. Section 3 introduces the proposed method. Section 4 presents the experimental results and the conclusion is given in Section 5.

2 Low-Rank Representation

2.1 Algorithm Description

As proposed in [9], for a set of data vectors $X = [x_1, x_2, \dots, x_n]$ (each column is a sample) in R^D , the decomposition model that approximates matrix X can be characterized:

$$\begin{aligned} \min_Z \quad & \|Z\|_* + \lambda \|E\|_{2,1}, \\ \text{s.t.}, \quad & X = AZ + E, \end{aligned} \tag{1}$$

where $Z = [z_1, z_2, \dots, z_n]$ is the coefficient matrix with each z_i being the representation of x_i , $A = [a_1, a_2, \dots, a_m]$ is the overcomplete dictionary which can represent x_i by the linear combination of its basis, $\|E\|_{2,1} = \sum_{j=1}^n \sqrt{\sum_{i=1}^m ([E]_{ij})^2}$ is called as the $\ell_{2,1}$ -norm, and the parameter $\lambda > 0$ is used to balance the effects of the two parts. Here, $\|\cdot\|_*$ denotes the *nuclear norm* [12] of a matrix, *i.e.*, the sum of the singular values of the matrix.

Using the data X itself as the dictionary, Eq. (1) can be converted into:

$$\begin{aligned} \min_Z \quad & \|Z\|_* + \lambda \|E\|_{2,1}, \\ \text{s.t.}, \quad & X = XZ + E, \end{aligned} \tag{2}$$

where XZ is the low-rank matrix, and E is the sparse matrix.

2.2 Solution to the Optimization Problem

For Eq. (2), it can be viewed as the following equivalent problem:

$$\begin{aligned} \min_{Z,E,J} \quad & \|J\|_* + \lambda \|E\|_{2,1}, \\ \text{s.t.}, \quad & X = XZ + E, \\ & Z = J, \end{aligned} \tag{3}$$

which can be solved by Augmented Lagrange Multiplier (ALM) method [13]:

$$\begin{aligned} \min_{Z,E,J} \quad & \|J\|_* + \lambda \|E\|_{2,1} + \\ & \text{tr}[Y_1'(X - XZ - E)] + \text{tr}[Y_2'(Z - J)] + \\ & \frac{\mu}{2} (\|X - XZ - E\|_F^2 + \|Z - J\|_F^2), \end{aligned} \tag{4}$$

where Y_1 and Y_2 are Lagrange multipliers and $\mu > 0$ is a penalty parameter. We outline the inexact ALM in Algorithm 1. Note that Steps 1 and 3 of the algorithm are convex problems. However, both of them have closed-form solutions. Particularly Step 1 is solved via singular value thresholding operator [14], and Step 3 can be solved by the following lemma [9]:

Lemma Let $Q = [q_1, q_2, \dots, q_i, \dots]$ be a given matrix and $\|\cdot\|_F$ be the Frobenius norm. If the optimal solution of

$$\min_w \lambda \|W\|_{2,1} + \frac{1}{2} \|W - Q\|_F^2$$

is W^* , then the i -th column of W^* is

$$W^*(:,i) = \begin{cases} \frac{\|q_i\| - \lambda}{\|q_i\|} q_i, & \text{if } \lambda < \|q_i\| \\ 0, & \text{otherwise.} \end{cases}$$

Algorithm 1. Solving Eq.(2) by Inexact ALM

Input: data matrix X , parameter λ

Initialize: $Z = J = 0$, $E = 0$, $Y_1 = 0$, $Y_2 = 0$, $\mu = 10^{-6}$, $max_u = 10^{10}$, $\rho = 1.1$, $\varepsilon = 10^{-8}$.

while not converged **do**

1. fix the others and update J by

$$J = \arg \min \frac{1}{\mu} \|J\|_* + \frac{1}{2} \|J - (Z + Y_2 / \mu)\|_F^2$$

2. fix the others and update Z by

$$Z = (I + X'X)^{-1} (X'X - X'E + J + (X'Y_1 - Y_2) / \mu)$$

3. fix the others and update E by

$$E = \arg \min \frac{\lambda}{\mu} \|E\|_{2,1} + \frac{1}{2} \|E - (X - XZ + Y_1 / \mu)\|_F^2$$

4. update the multipliers

$$Y_1 = Y_1 + \mu (X - XZ - E)$$

$$Y_2 = Y_2 + \mu (Z - J)$$

5. update the parameter μ by $\mu = \min(\rho\mu, max_u)$

6. check the convergence conditions

$$\|X - XZ - E\|_\infty < \varepsilon \text{ and } \|Z - J\|_\infty < \varepsilon$$

end while

3 The Proposed Method

Given a pair of multi-temporal remote sensing images X_1 and X_2 , we denote it as $X = [x_1, x_2]$. Both x_1 and x_2 are column vectors which are reshaped by X_1 and X_2 . For change detection, the majority of the images X_1 and X_2 is usually the unchanged areas, and thus the data of unchanged areas is low-rank. Moreover, the data of changed areas can be viewed as sparse. Therefore, we use LRR to decompose

X according to Eq. (2). It should be noted that the low-rank part XZ corresponds to the unchanged areas and the sparse part E corresponds to the changed areas. In this way, we can generate the difference image from the sparse part E .

However, each method has its own applicability in applications, and thus a single method cannot be suitable for all types of images. In order to achieve the complementary advantages of the algorithms, we propose to use low-rank representation again to obtain low rank part with difference images obtained by different methods.

In this paper, we firstly apply mean filter for denoising, and then use the subtraction and logarithm ratio operators to obtain two initial difference images, as shown in Eqs. (5) and (6):

$$D_s = |X_1 - X_2| \quad (5)$$

$$D_l = \left| \log \frac{X_2 + 1}{X_1 + 1} \right| = |\log(X_2 + 1) - \log(X_1 + 1)| \quad (6)$$

where D_s represents the difference image by the subtraction operator, and D_l represents the difference image by the logarithm operator. At the same time, we can obtain another initial difference image generated by LRR. After that, we utilize LRR to obtain the low rank part of these three difference images, which can reflect the common characteristics and thus can be viewed as the final difference image under three different methods. Finally, k -means is applied for post-processing. The flow chart of our method is illustrated in Fig. 1.

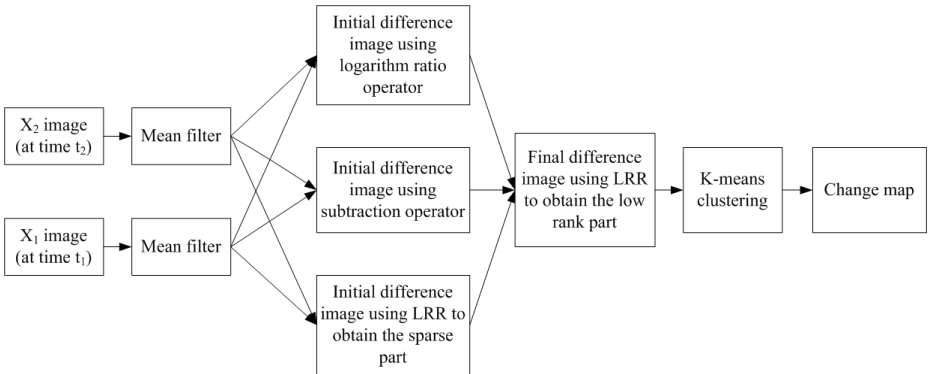


Fig. 1. Flow chart of our method

4 Experimental Results

4.1 Description of Datasets

In order to assess the effectiveness of our method, we carried out the experiments on two real multi-temporal datasets [15].

The first dataset represents a section (512×512 pixels) of two remote sensing images. Fig. 2(a) and (b) are acquired by Band 4 of the Landsat Enhanced Thematic Mapper Plus (ETM+) sensor of the Landsat-7 satellite in an area of Mexico in April 2000 and May 2002. A reference map was manually defined as shown in Fig. 2(c).

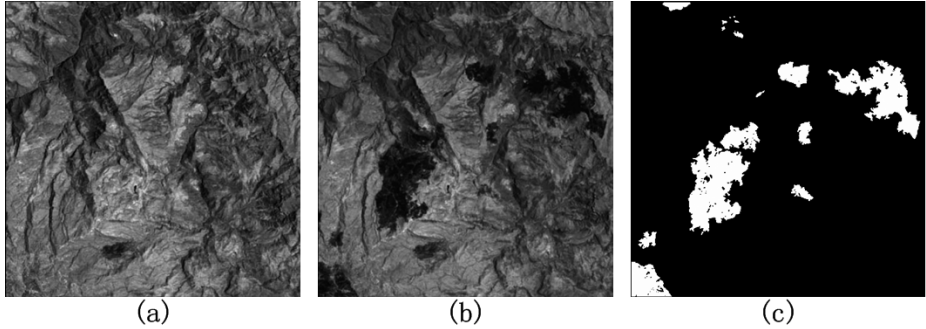


Fig. 2. Multi-images relating to the area of Mexico. (a) The image acquired in April 2000, (b) The image acquired in May 2002, (c) The reference map.

The second dataset represents a section (412×300 pixels) of two remote sensing images. Figs. 3(a) and (b) are acquired by Band 4 of the Landsat Thematic Mapper (TM) sensor of the Landsat-5 satellite on the Lake Mulargia on the Island of Sardinia (Italy) in September 1995 and July 1996. A reference map was manually defined as shown in Fig. 3(c).

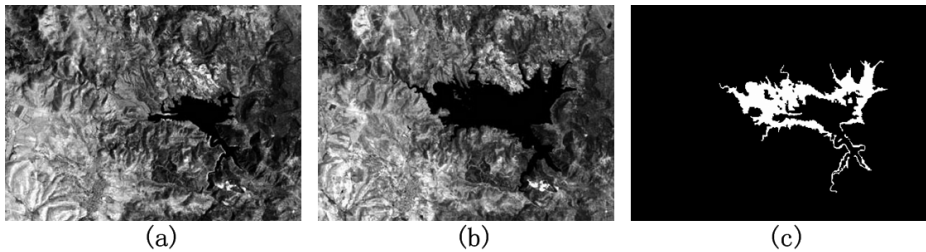


Fig. 3. Multi-images relating to Sardinia. (a) The image acquired in September 1995, (b) The image acquired in July 1996, (c) The reference map.

4.2 Results and Analysis

In this section, we compare our method with three widely used methods (*i.e.*, Nonsub-sampled Contourlet Transform (NCT) [16], Dual-Tree Complex Wavelet Transform (DTC) [17] and Undecimated Discrete Wavelet Transform (UDWT) [18]).

In our experiments, we set the parameter λ of LRR to 26×10^4 and 12×10^4 for the first and second experiments, respectively. The final results of change detection under different methods are shown in Figs. 4 and 5. It is obvious that DTC and UDWT have a lot of isolate pixels. NCT leads to the loss of the large amount of edge

information. In contrast with other methods, our method shows good performance in the detection of fine edge part and has less isolate pixels. It should be pointed out that the result obtained by the proposed method is very close to the reference map, from the visual analysis of Figs. 4 and 5.

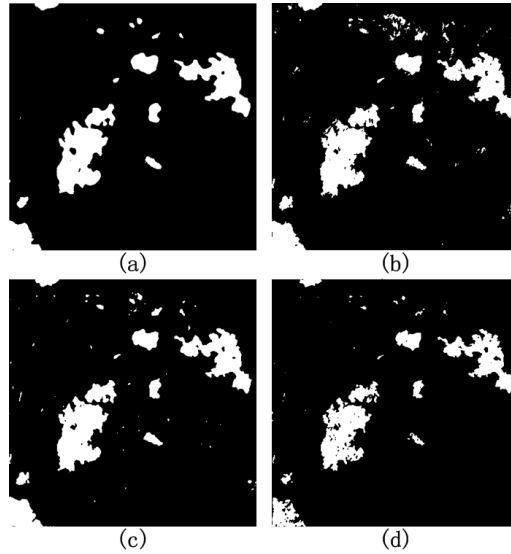


Fig. 4. Change detection results of different methods on the Mexico dataset. (a) NTC, (b) DTC, (c) UDWT, (d) our method.

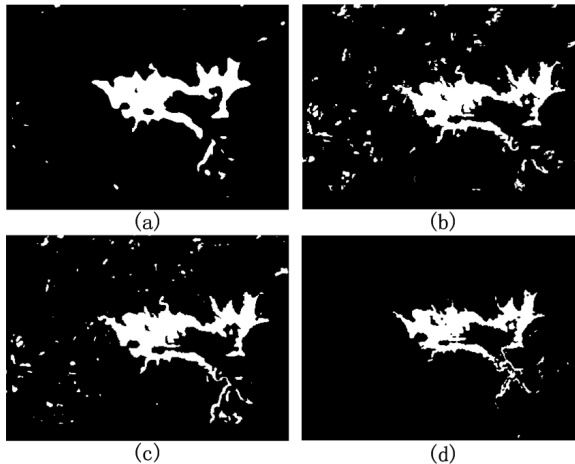


Fig. 5. Change detection results of different methods on the Sardinia dataset. (a) NTC, (b) DTC, (c) UDWT, (d) our method.

Tables 1 and 2 show the quantitative results for the two datasets respectively. As can be seen, our method has the least number of false alarms and the most number of

missed alarms compared with other methods. But, our method has the least number of overall error than other methods. It may be due to that our method uses LRR to fuse three initial difference images, which can reflect the common characteristics of them.

Table 1. Comparison of detection results on the Mexico dataset (in number of pixels)

Methods	False alarms	Missed alarms	Overall Error
NCT	2824	1847	4671
DTC	3698	834	4532
UDWT	2855	1834	4689
Ours	855	3622	4477

Table 2. Comparison of detection results on the Sardinia dataset (in number of pixels)

Methods	False alarms	Missed alarms	Overall Error
NCT	3005	583	3588
DTC	3821	400	4221
UDWT	2939	370	3309
Ours	1479	886	2365

5 Conclusion

In this paper, we have proposed a novel change detection method for remote sensing image, which is based on low-rank representation. It uses the subtraction operator, logarithm ratio operator and LRR to obtain three different difference images, and then combines them to extract commonness parts by LRR. Finally k -means is used to acquire the final result of change map. By using LRR, our method not only reveals the specific characteristics of image content, but also achieves the complementary advantages of different kinds of change detection methods. Experimental results show that our method can effectively obtain difference images and simultaneously outperform state-of-the-art methods.

Acknowledgement. This work was supported by the National Natural Science Foundation of China (No. 61071137, 61371134), and the 973 Program of China (Project No. 2010CB327900).

References

1. Bruzzone, L., Serpico, S.B.: An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 35(4), 858–867 (1997)
2. Hame, T., Heiler, I., San Miguel-Ayanz, J.: An unsupervised change detection and recognition system for forestry. *International Journal of Remote Sensing* 19(6), 1079–1099 (1998)

3. Di Martino, G., Iodice, A., Riccio, D., Ruello, G.: A novel approach for disaster monitoring: Fractal models and tools. *IEEE Transactions on Geoscience and Remote Sensing* 45(6), 1559–1570 (2007)
4. Ridd, M.K., Liu, J.: A comparison of four algorithms for change detection in an urban environment. *Remote Sensing of Environment* 63(2), 95–100 (1998)
5. Singh, A.: Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing* 10(6), 989–1003 (1989)
6. Bruzzone, L., Prieto, D.F.: Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Geoscience and Remote Sensing* 38(3), 1171–1182 (2000)
7. Chang, C.I., Wang, S.: Constrained band selection for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* 44(6), 1575–1585 (2006)
8. Townshend, J.R.G., Justice, C.O.: Spatial variability of images and the monitoring of changes in the normalized difference vegetation index. *International Journal of Remote Sensing* 16(12), 2187–2195 (1995)
9. Liu, G., Lin, Z., Yu, Y.: Robust subspace segmentation by low-rank representation. In: 27th International Conference on Machine Learning, pp. 663–670. ICML Press, Haifa (2010)
10. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(2), 210–227 (2009)
11. Shi, J., Jiang, Z.G., Feng, H., Ma, Y.B.: Sparse coding-based topic model for remote sensing image segmentation. In: *IEEE Geoscience and Remote Sensing Symposium*, pp. 4122–4125. IEEE Press, Melbourne (2013)
12. Fazel, M.: Matrix rank minimization with applications. PhD thesis. Stanford University (2002)
13. Lin, Z., Chen, M., Ma, Y.: The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv preprint arXiv: 1009.5055 (2010)
14. Cai, J.F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization* 20(4), 1956–1982 (2010)
15. Ghosh, S., Bruzzone, L., Patra, S., Bovolo, F., Ghosh, A.: A context-sensitive technique for unsupervised change detection based on Hopfield-type neural networks. *IEEE Transactions on Geoscience and Remote Sensing* 45(3), 778–789 (2007)
16. Li, S., Fang, L., Yin, H.: Multitemporal Image Change Detection Using a Detail-Enhancing Approach With Nonsubsampled Contourlet Transform. *IEEE Geoscience and Remote Sensing Letters* 9(5), 836–840 (2012)
17. Celik, T., Ma, K.K.: Unsupervised change detection for satellite images using dual-tree complex wavelet transform. *IEEE Transactions on Geoscience and Remote Sensing* 48(3), 1199–1210 (2010)
18. Celik, T.: Multiscale change detection in multitemporal satellite images. *IEEE Geoscience and Remote Sensing Letters* 6(4), 820–824 (2009)

Two New Methods for Locating Outer Edge of Iris*

Yujie Liu and Hong Tian

Institute of Software of Dalian Jiaotong University, Dalian, China
amikejingling@163.com,
tianhw@263.net

Abstract. The two new methods for locating the outer edge of iris were presented by researching of iris images. The first is to locate the inner edge of iris by using canny operator and Hough transform method. The selection of the appropriate threshold is based on the diagram which observes the color distribution of the iris image; the second is that the circle integral and linear segment methods to locate the outer edge of iris are presented. We would like to indicate that the experiment results show that both of the two methods have the following qualities: rapidity, availability, real time calculations and accuracy in locating the iris outer edge.

Keywords: Iris localization, circle integral and linear segment.

1 Introduction

With the growth of the economy and the development of the science and technology, information security has become one of the most pertinent questions. Biometric character identification technology has played a vital role for the improvement of the information security. In numerous biometric character identification technologies, iris recognition technology has attracted a growing number of mathematicians and scientists to explore since its special uniqueness, stability, detect-ability, security, which is widely regarded as the most promising biometric technologies in 21st century [1].

Iris image segmentation is one of the most important steps in the process of the iris recognition, and the iris localization is a very critical step for the iris image segmentation. The quality of the iris localization directly affects the iris recognition. Different approaches have already been reported in the literature to locate the iris region. John Daugma [2] in 1993 proposed the first efficient iris location method. He located the iris boundaries using a relatively time-consuming differential operator. Wildes [3] then used the gradient criterion and circular Hough transform to locate the iris which votes the edge information in binary image and determines the iris edge parameters according to the votes. Although both of the above methods can locate relatively stable iris edge, but both of them are time consuming and cannot be applied in real-time system. Zhang Zaifeng [5] put forward a kind of method in 2009 which was based on

* The project is supported by National Natural Science Foundation of China (61074029) and Natural Science Foundation of Liaoning Province (20102014).

the combination of the threshold segmentation and the radial gradient local maximum values to locate the iris by using 23ms. In 2011, He Xiaofu [4], Tian Xuzi, Zhang Yuan and Huang Liyu [6] improved the method of Hough transform to locate the iris edge and the time they locate iris were 0.56s and 2.5s. The following methods also cannot be used in real-time system either.

First, we use the canny operator and the Hough transforms to extract iris inner edge. The threshold of Hough transform is determined by the histogram method. After that, we observed the iris images and segmented sub-images based on the original images which regard the pupil center as their center and the size of these sub-images is 240*220. These images contain the entire iris image and decrease the number of pixels. Thus the speed of locating the outer edge of iris images has been increased. Finally, aiming at the problem of the blurring outer edge of iris, we present the circle integral and linear segment methods to locate iris outer edge.

2 The Localization of Inner Edge of Iris

The iris inner edge is just the edge between the pupil and iris which is clear. Moreover the inner edge can be approximately considered as a circle. And the operation of smooth can be used to reduce noise. Therefore, we use the method of Hough circles transform to detect and locate the pupil region.

Hough transform method is a kind of gradient method which can solve the circle transform problem. First of all pupil area can be distinguished by applying a threshold step and the method of diagram can help to choose an exact threshold which are used in the image binary progress. Canny edge detector can be used at next step to get edge image. At last using the Hough method to estimate the pupil center and radius on the edge image (x_j, y_j) . The work is described in the following equations.

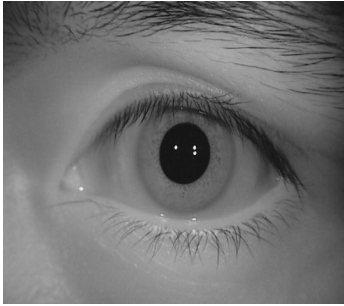
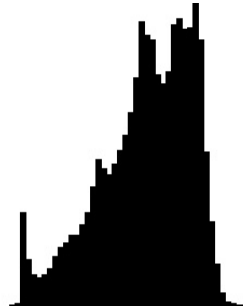
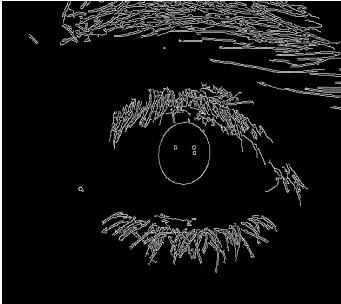
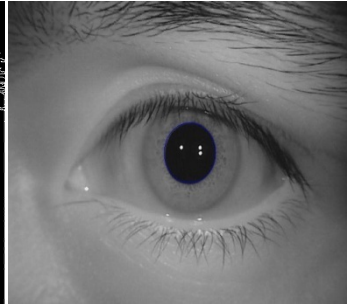
$$H(x_c, y_c, r) = \sum_{j=1}^n h(x_j, y_j, x_c, y_c, r) \quad (1)$$

Where x_j and y_j present the coordinates of the circle centers, r presents radius of every circle which has been detected. These three parameters constitute some parameter sets and H presents an accumulator to choose candidate center by voting that parameter sets. Selecting centers which meet the accumulator condition and sorting these centers along with the cumulative value of votes in descending order.

$$h(x_j, y_j, x_c, y_c, r) = \begin{cases} 1, & g(x_j, y_j, x_c, y_c, r) = 0 \\ 0, & g(x_j, y_j, x_c, y_c, r) \neq 0 \end{cases} \quad (2)$$

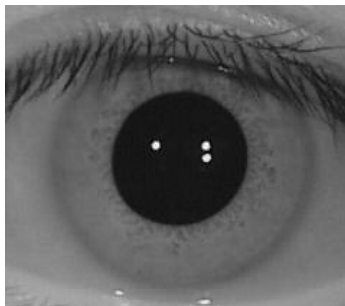
Where $g(x_j, y_j, x_c, y_c, r) = (x_j - x_c)^2 + (y_j - y_c)^2 - r^2$ denotes the voting function which satisfies the parameter equation of circle. Voting the parameter group (x_c, y_c, r) when the edge point (x_j, y_j) is on the circle which consists of the parameter group (x_j, y_j, x_c, y_c, r) . According to the method, we can get some circles of iris image except the pupil's edge. So, we stipulate the scope of the size of the pupil and the distance between the pupil center and the center of the image.

An example of this process on an eye image is shown from Fig. 1 to Fig.4.

**Fig. 1.** Input image**Fig. 2.** Gray histogram of the image**Fig. 3.** Edge image**Fig. 4.** Inner edge of iris image

3 Image Segmentation

Iris is a small part of the whole eye image, and the redundant part greatly increases the workload of the computer and the time of locating iris. Furthermore, a series of factors (i.e., eyebrows, eyelids) may affect the location of outer edge of iris. So we intercept a sub-image from input image (the size of which size 240*220).

**Fig. 5.** Sub-image

This method can be prominent the iris image without the whole eyebrows and some parts of eyelids. The sub-image had narrowed down to the 1/6 size of the original image.

4 The Localization of Outer Edge of Iris

The difference between iris and pupil pixels is not clear when the iris images in gray-scale map. Marginalized by canny operator, only the parts of pupil is obvious. No matter how many times one needs to change the threshold, the edge of iris cannot be observed without any other operation.

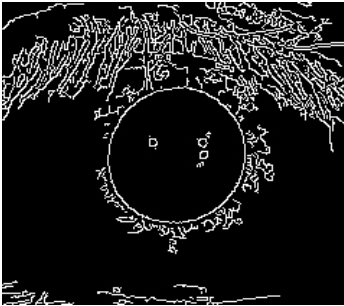


Fig. 6. Edge image (a)

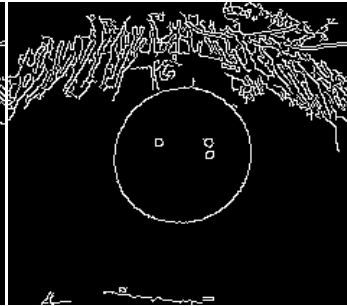


Fig. 7. Edge image(b)

Two of the above pictures are edge images after the use of canny edge detector. Figure 6 uses a smaller threshold value than figure 7, but both of them cannot show any part of the outer edge. Therefore, we cannot detect the outer edge by using the simple method and this texture offer two methods which are circle integral and linear segment to solve the problem.

4.1 Circle Integral

According to the feature that the iris and the pupil can be regarded as two concentric circles, the center and radius of pupil can be used to identify the outer edge of the iris. Searching in the $b \times b$ neighborhood of pupil center, one realizes that the scope of search radius is between $\min(r)$ and $\max(R)$. This process is described below:

$$\text{SumPixel}_{(i \times j, R_0 - r)} = \sum_{m=1}^{220} \sum_{n=1}^{240} P_{mn} \quad (1 < i, j < 6) \tag{3}$$

$$\min(r) < (x_{mn} - x_{ij})^2 + (y_{mn} - y_{ij})^2 < \max(R) \tag{4}$$

Where (x_{ij}, y_{ij}) denotes the coordinate of the neighborhood of pupil center, and (x_{mn}, y_{mn}) denotes the coordinate of the image. Where m and n denote the satisfied following formula.

Calculating the absolute distance of SumPixel:

$$|\text{SumPixel}_{i+1} - \text{SumPixel}_i| \tag{5}$$

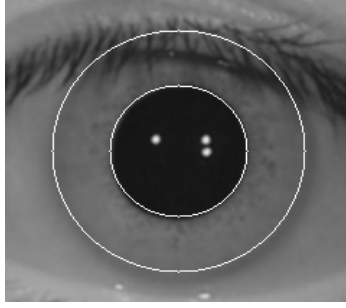


Fig. 8. The image of outer edge of iris

4.2 Linear Segment

The sub-image also has some noises (i.e. eyelashes) which hinder the extraction of the limbic edge. So a quarter of the up and down part has been got rid of the iris image, only to search the middle section and look for external edge.

Here are the steps to follow:

- (1) Segment the sub-image, and the size of the rest image is $240 * 220$.
- (2) Get each line elements first derivative of the rest of the image respectively.
- (3) Divide the elements of each row into three parts: the pupil and the left part and right part of the pupil. Scanning the left part and the right part of the pupil. Because the pixel values of iris are smaller than the part of sclera, the max value is on the left and the min value is on the right. Recording the coordinates of the max and min values. Then some points of the outer edge are found.
- (4) Use the method of fitting to get the outer edge of the iris.

The result of the linear segment process is shown in Fig.9

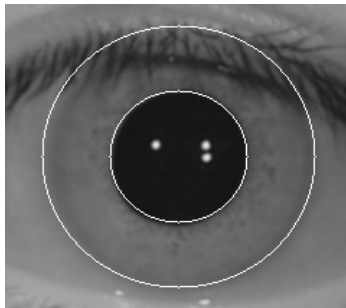


Fig. 9. Outer edge of iris by used the method of linear segment

5 Contrast Experiment

The iris database of CASIA1.0 from Chinese Academy of Sciences Institute of Automation has been used to verify the accuracy and real-time of the two methods. Results of the different methods of the Hough, calculus, circle integral and linear segment are shown in Fig. 10.

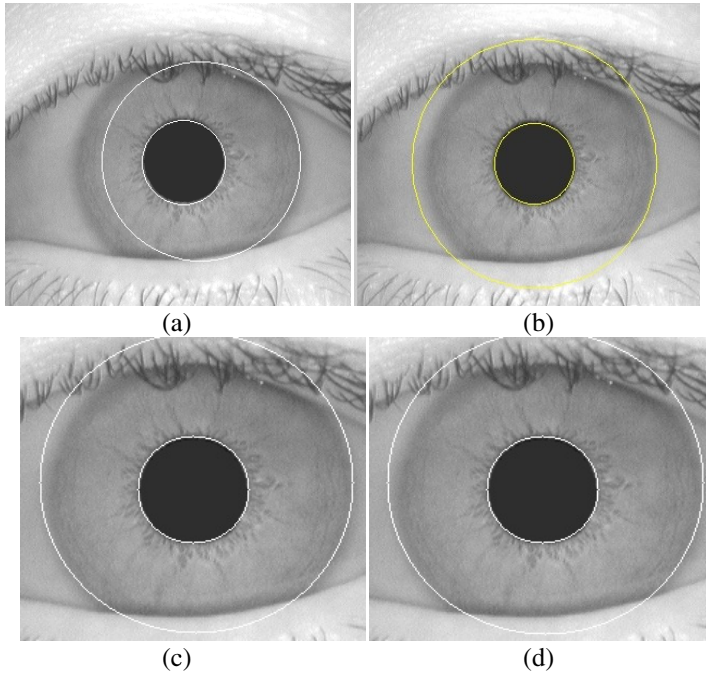


Fig. 10. a) Outer edge of iris by used the method of Hough, b) Outer edge of iris by used the method of calculus c) Outer edge of iris by used the method of circle integral d) Outer edge of iris by used the method of linear segment

Table 1. The time of four methods

Location method	time (s)
Hough	0.75
calculus	0.437
circle integral	0.015
linear segment	0.008

The average times by using these four methods are 0.76s, 0.42s, 0.01s and 0.003s. The two methods of this essay have already increased the speed.

6 Conclusions

The results of implementing the location methods by OpenCV and c on a computer with 1.73GHZ processor and 1 GB RAM have been reported. Algorithm is tested on CASIA1.0 database consisting of 756 images of 108 persons. The present study has proposed two algorithms for the iris localization which robust against disturbances of eyelids and eyelash and unwanted edges.

Circle integral method search the image around pupil center and the search radius is the scope of $[\min (r), \text{Max} (R)]$ which can decrease the amount of background calculation. The second method removes the upper and lower part of the image; search the left and right part of the middle layer to get the points of the outer edge of iris. Circle fitting can help to find the iris's center and radius in the image. Finally, these two methods can be used to improve the accuracy and speed in locating the iris.

References

1. Wan, G.: Application and development of iris recognition technology. *Marine Electric Technology* 28(5), 308–311 (2008)
2. Daugman, J.G.: High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(11), 1148–1161 (1993)
3. Richard, P.W.: Iris recognition: An emerging biometric technology. *Proceedings of IEEE* 85(9), 1348–1363 (1997)
4. He, X.: The key technology of live iris recognition research. Shanghai Jiaotong University (2007)
5. Zhang, Z.: Iris recognition algorithm research and system implementation. Lanzhou University (2009)
6. Tian, X., Zhang, Y., Huang, L.: Combined with morphology of Hough transform iris localization algorithm. *Computer Engineering and Application* (2012)

Effects of the Gridding to Numerical Simulation of the Armour-Piercing Warhead Penetration through the Steel Target

Jun-qing Huang, Ya-long Fan Rui Ma, and Wei Shao

Department of Equipment Command and Administration, Academy of Armored Force
Engineering, Beijing, China
tigerhjq@126.com

Abstract. paper studies the influencing rule of gridding definition for calculating result in numerical simulation of the armour-piercing warhead penetration through the steel target, so we can get the reasonable scope of grid size. Paper has used the adopted explicit dynamic analyzing program AUTODYN to simulate the process of the armour-piercing penetration through the steel target. Based on the simulation, we get the penetration deepness, penetration overload and damage area of the target in conditions of different grid size. In order to get more reasonable result, the scope of grid size is equaled to about 5.0mm by contrasting and analyzing.

Keywords: Armour-piercing, penetration, numerical simulation and grid size.

1 Introduction

With the rapid development of computer technology, the application of numerical method becomes possibly for the simulation process of the armor-piercing warhead penetration target^[1]. As the base of numerical simulation, partition of grid size had direct effect to numerical simulation result. For dynamic load, lots of elements were added to material model, such as the strain rate, so the dependencies of grid is more important. Since the effect of strain rate is lying on partition of grid size, dynamic strength also lies on the partition of grid size. If a big grid size was defined, the numerical simulation result would not be reasonable; and if small grid size was defined, the numerical simulation result would be reasonable, but the numerical simulation process would need much calculative time. For solving the problem of dependencies on grid in the numerical calculation, the common way is averaging the grid and calculate coarse grid model and subtle grid model respectively. If the numerical result was basically identical by adopting two different partition gridding projects, it showed that the grid size could be used to numerical calculation. In this essay, the explicit dynamic analyzing program AUTODYN based on finite element was adopted to simulate the process of the armor-piercing warhead penetration through the steel target which was defined by different grid size, and the calculative result was analyzed in order to get reasonable size.

2 Numerical Simulation Model

The following is an example. The armor-piercing warhead lateral section diameter 2.7cm, the length as 51.2cm, peripheral fixed target thickness is 22cm; the warhead and the target surface normal direction had 90.0° angle; the warhead speed was the 1494m/s. Fig.1 had given the warhead finite element discretization model structure and the grid schematic drawing (considered computation time and cycle, carries on model simplification, for example, for armor-piercing warhead was used to penetrate the target, so the warhead was retained, and the sabot, base fuze etc. wear neglected) .

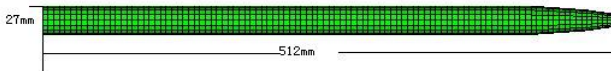


Fig. 1. ModelGrid Schematic Drawing

2.1 Material Model

Material model mainly included constitutive model, state equation and strength model^[2]. Lagrange calculative way was adopted for only discussing the character of calculative result lying on gridding. Material models of the armor-piercing warhead and target wear from AUTODYN material models database. The mainly material models of the armor-piercing warhead and target wear in Table 1.

Table 1. The Mainly Material Models of The Armor Piercing Warhead and Target

part name	material	state equation	strength model
warhead	tungsten alloy	Grüneisen	Johnson-Cook
target	steel	Shock	Johnson-Cook

2.2 Gridding Partition

The Lagrange way was used to the armor-piercing warhead and target numerical model. For the armor-piercing warhead perpendicularity penetration the target and the model having the character of axis symmetry, so 1/2 model was built. For studying the effect of the target grid size to penetration, the gridding shape was square, and the grid size was deal with by dimension. K was defined for gridding side length, unit was millimeter. The K value was respectively equaled to 20mm, 10mm, 5mm, and 4mm. Fig.2 showed the armor-piercing warhead and target numerical model built by different gridding side length. Table 2 showed the relation of gridding side length and element number.

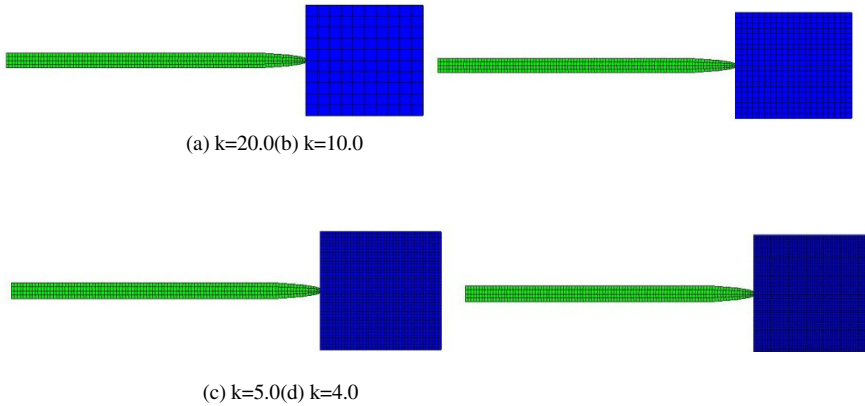


Fig. 2. The Armor-Piercing Warhead and Target Model of Different k Value

Table 2. The Target Gridding Partition

Partition project	Element length	Element number
a	20.0	792
b	10.0	5082
c	5.0	36162
d	4.0	68952

3 Numerical Simulation Results and Analyses

By numerical simulation of the armor-piercing warhead penetration the steel target, influent rule of different k value gridding side length to the armor-piercing warhead penetration velocity and deepness, damage area of the target and numerical simulation efficiency was acquired

3.1 Effects of Gridding to the Armor-Piercing Warhead Penetration Velocity and Impulse

Fig.3 showed the armor-piercing warhead velocity along with time change course curve in different k value.

Fig.4 showed the armor-piercing warhead impulse along with time change course curve in different k value.

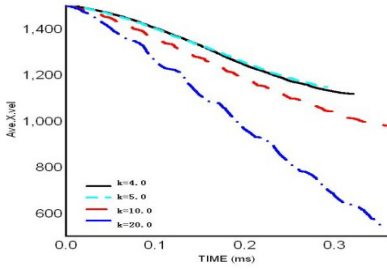


Fig. 3. WarheadVelocityalong with Time Change Course Curve in Different k Value

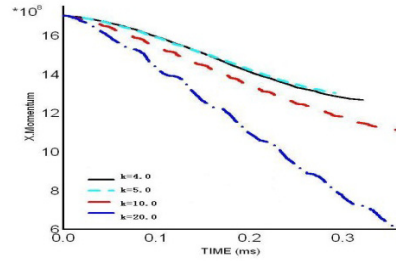


Fig. 4. Warheadimpuse along with time change course curve in different k value

The numerical simulation result showed: the velocity and impulse of the armor-piercing warhead penetration the target wear gradually decrease because it was underwent resistance of the target in the armor-piercing warhead penetration the target^[3]. The decrease extent of the velocity and impulse was continually increased. As the k value was equal to 4.0 and 5.0, the calculative result of velocity and impulse was basically consistent, the most error was about 0.5% and it caught be neglected. As the k value was equal to 10.0, the velocity and impulse change trend wear basically same. Compared with the k value be equal to 4.0 and 5.0, the numerical result was basically consistent. But when the k value was equal to 20, the change of the armor-piercing warhead velocity and impulse along with time was gradually decrease, and the numerical result basically consistent comparing with the k value be less than 10.0. It was showed by theory and practice that the change of the armor-piercing warhead penetration velocity and impulse was rather tempered because in the press of the armor-piercing warhead penetration the target, the condition was built for the armor-piercing warhead penetration,it included relative steady high pressure, high strain and high distortion speed state in the target. So, the change of warhead velocity and impulse along with time was rather reasonable when the k value was equal to 4.0,5.0 and 10.0.

3.2 Effect of Gridding to Damage Area of the Target

The material model of steel included material damage description, Fig.5 showed the target damage state in different k value. Table.3 showed calculative result of the target damage area.

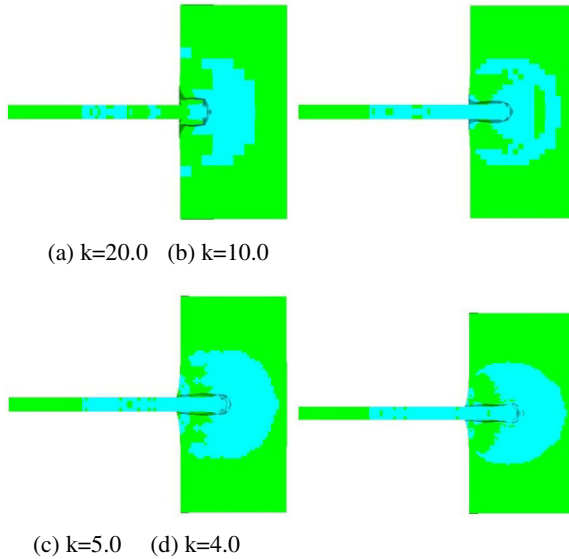


Fig. 5. The Target Damage State in Different k Value

Table 3. Calculative Result of the Target Damage Area in Different k Value

k	20.0	10.0	5.0	4.0
Damage area diameter	230.4	210.8	215.5	209.3

Table.3 showed that effect of the target damage area was small in different k value, but Fig.7 showed that the target damage area shape was affected in different k value. As the k value was equal to 4.0 and 5.0, the plastic strain of the target damage area was rather serial, and the brim of the target damage area was rather clear^[4]. As the k value was less than 4.0, the forehead of target had obvious bulgy phenomena. So, for acquiring better image of the target damage area, the k value ought to be equalued to less than 20.0.

3.3 Effect of Gridding to Numerical Simulation Efficiency

Table.4 showed calculative times of the computer when the depth of the armor-piercing warhead penetration the target arriving to 50mm, in computer environment of the Intel(R) Core(TM)2 Duo CPU E7400 2.80GHz 2.79GHz,2.0GB memory.

Table 4. Calculative Times in Different k Value

k	20.0	10.0	5.0	4.0
Calculative time	10	15	37	53

Table.4 showed that the computer time was continually increase along with decrease of the gridding size, and the numerical calculating needed more time if gridding size was more small. The computer time difference had more than 5 times by comparing with the k value be equal to 4.0 and 5.0.

4 Conclusion

The Lagrange methodis used to the armor-piercing warhead and the target numerical model. For choosing gridding size, in the condition of the armor-piercing warhead strength be enough (having no obvious distortion), the warhead diameter direction at least has 5 gridding, and gridding size of the steel target was confirmed base on numerical simulation effect. For acquiring perfect numerical simulation result, clear shape of damage area and reasonable calculative time, the k value ought to be equaled to about 5.0.

References

1. Backman, M.E., Goldsmith, W.: The mechanics of penetration of projectiles into targets. *International Journal of Engineering Science* 16(1), 1–108 (1998)
2. Wu, Y.: Ax symmetric Penetration of RHA steel targets by cylindrical tubes. In: *Proceedings of the 1995 International Conference on Metallurgical and Materials Applications of Shock-Wave and High-Strain-Rate Phenomena*, pp. 337–344 (1995)
3. Roessing, K.M., Mason, J.J.: Adiabatic shear localization in the dynamic punch test, part:numerical simulations. *Int. J. Plasticity* 15, 263–283 (1999)
4. Ramesh, K.T.: Localization in Tungsten Heavy Alloys Subjected to Shearing Deformations Under Superimposed High Pressures. *Metal Powder Industries Federation*, 3–9 (1995)

Author Index

- Bai, Suqin 213
Bao-feng, Tong 315

Changwei, Zheng 257
Chen, Yisong 153
Cheng, Yan 336
Cong, Li 181
Cong, Lin 137
Cui, Jian 175

Dan, Wang 29
Dong, Jing 305

Fan, Xu-dong 224

Gang, Fu 181
Gu, Boyu 74
Guo, Feng 8
Guo, Wei 137

Hongrui, Zhao 181
Hou, Zhiqiang 83
Hu, Jie 48
Huang, Jun-qing 352
Huang, Xiyi 264
Huimin, Ma 91

Jiang, Zhiguo 327, 336
Jing, Junfeng 274
Jun, Du 239
Jun-yuan, Zhang 315

Li, Congli 1
Li, Guanghui 224, 264
Li, Pengfei 274
Liang, Qiang 224, 230, 264
Lin, Tao 48
Lin, Xiaozhu 120
Lin, Xuan 274
Liu, Guodong 67
Liu, Hao 175
Liu, Juhua 103
Liu, Ning 294
Liu, Xiaolin 40
Liu, Yang 153

Liu, Yujie 345
Liu, Zhanlin 120
Lu, Hesong 230, 264
Lu, Wenjun 1
Luo, Xiaonan 203
Luyao, Zhou 91

Ma, Ya-long Fan Rui 352
Ma, Yibing 327
Meng, Gang 336
Miao, Minjing 103
Minghui, Wang 57

Nan, Wang 315

Pang, Linbin 213

Qian, Qiang 213
Qiang, Liang 239
Qing, Xue 257

Ren, Pu 165
Renjie, Xu 239
Rui, Fan 239

Shao, Wei 224, 230, 264, 352
Shi, Jinlong 213
Shi, Jun 327, 336
Shi, Yongchang 1
Shi, Yue 110
Shi, Yuying 8
Song, Haodong 120
Su, Xinhua 8
Sun, Guoxia 284
Sun, Xiaoning 1

Tan, Lei 40
Tan, Tieniu 305
Tan, Yaxin 249
Tang, Zaijiang 230
Tian, Hong 345
Tian, Xiaohua 83
Tian, Yun 145

Wang, Guoping 153, 192
Wang, Jianyong 129
Wang, Liu 137

- Wang, Ruomei 203
Wang, Wei 48, 305
Wang, Weizhou 145
Wang, Wenjian 165
Wang, Xuemei 129
Wang, Zairan 305
Wang, Zhi 213
Wei, Xuefeng 145
Wei-guo, Liu 315
Wenchao, Xu 257
- Xi, Mingxiao 294
Xie, Yuting 48
Xie, Zhihua 67
Xingang, Peng 181
Xu, Chongbin 165
Xu, Dan 19
Xu, Haohua 249
Xu, Jing 8
Xu, Liang 145
Xu, Wanjun 83
Xu, Xiangzhong 249
Xue, Kaichuang 40
Xuejun, Li 57
- Yan, Yajing 294
Yang, Jiandong 249
- Yang, Tao 284
Yang, Xiaogang 129
Yi, Hanfei 103
Yin, Yabo 145
Yu, Wangsheng 83
Yuan, Yuan 103
- Zeng, Shan 203
Zhang, Chengfang 192
Zhang, Guodong 137
Zhang, Guoying 110
Zhang, Haopeng 336
Zhang, Lang 83
Zhang, Lei 274
Zhang, Qiang 74
Zhang, Zimin 175
Zhao, Aigang 129
Zhao, Yili 19
Zhao, Zhenbing 294
Zhao, Zhenhuan 74
Zheng, Yushan 327
Zhou, Fan 203
Zhou, Mingquan 145, 165
Zhou, Yi 19
Zhou, Ying 175
Zhou, Zhou 284